The Data-Acquisition System and new GPU-based High Level Trigger of the KOTO Experiment

Mario Gonzalez (Osaka University)
on Behalf of the KOTO Collaboration

July 19, 2024

Introduction: The KOTO experiment

- The KOTO experiment aims to measure the Branching ratio (BR) of $K_L \to \pi^0 \nu \bar{\nu}$
  - CP violating process
  - Theoretically predicted to be very small:

$$BR(K_L \to \pi^0 \nu \bar{\nu}) \sim 3\times10^{-11}$$

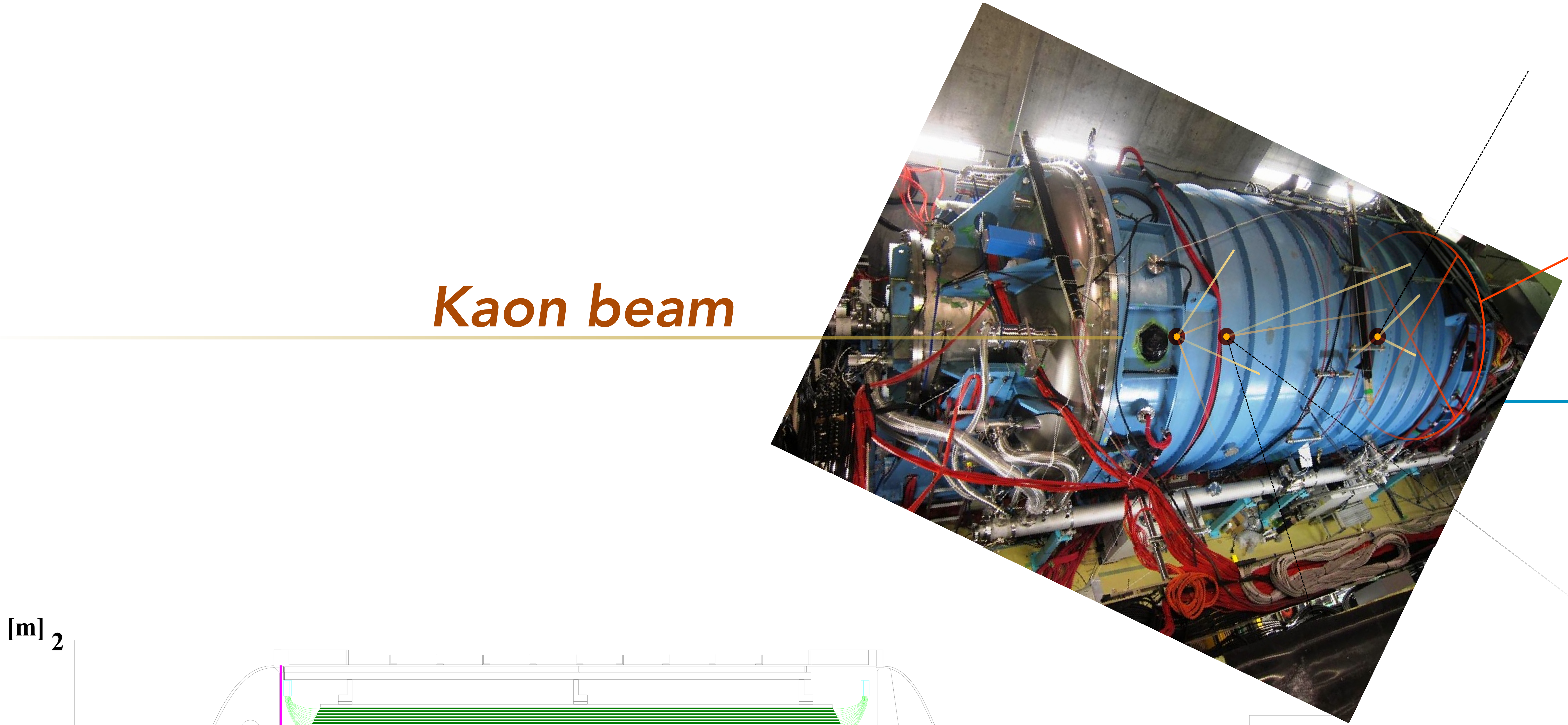
- Any deviation from the this value measured experimentally would indicate physics beyond the SM

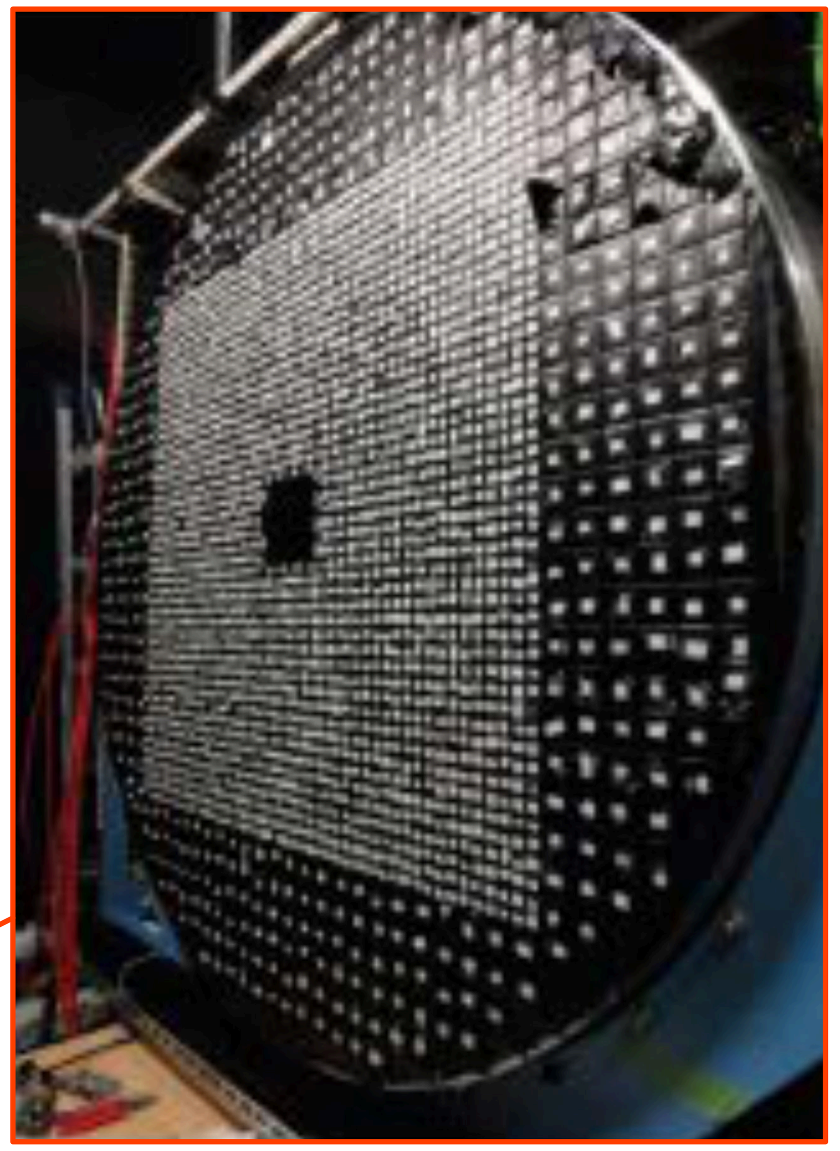
Experimental facility located at J-PARC

Data stored at the computing center at KEK

In the Ibaraki prefecture (Japan)

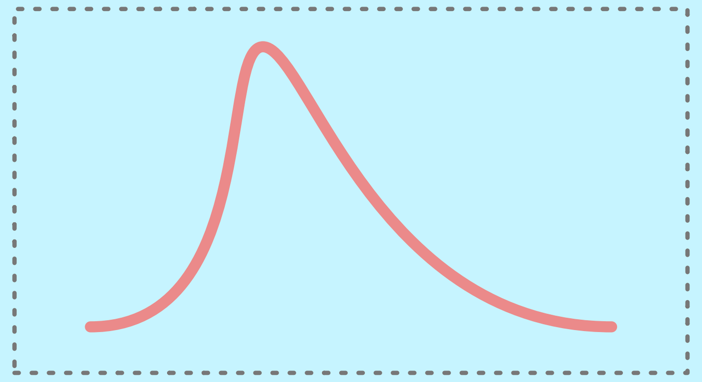
Introduction: The KOTO experiment

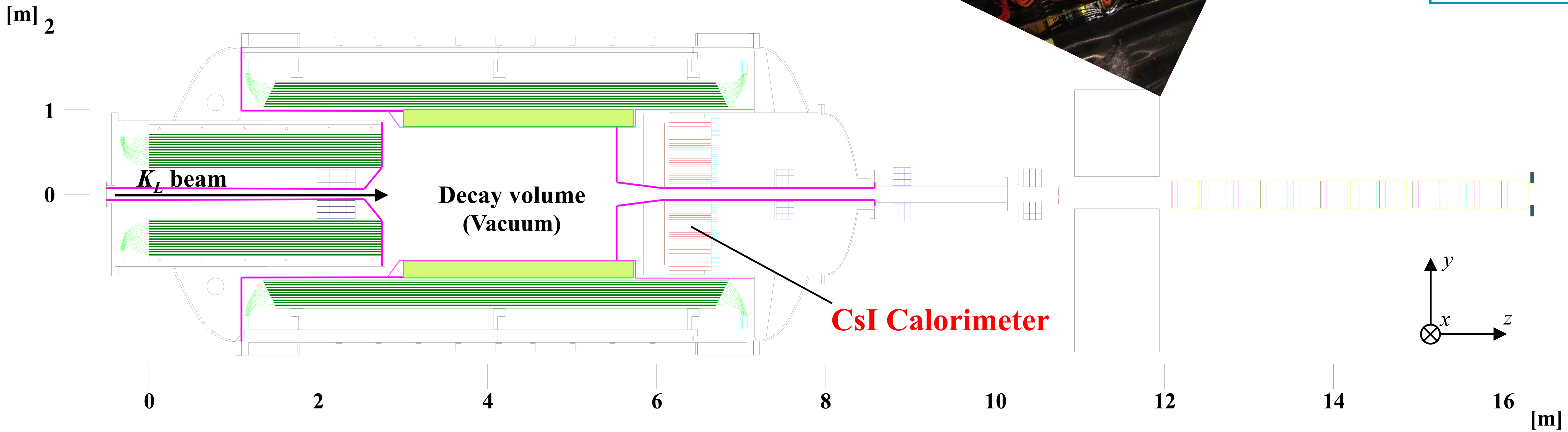




~4000 readout channels

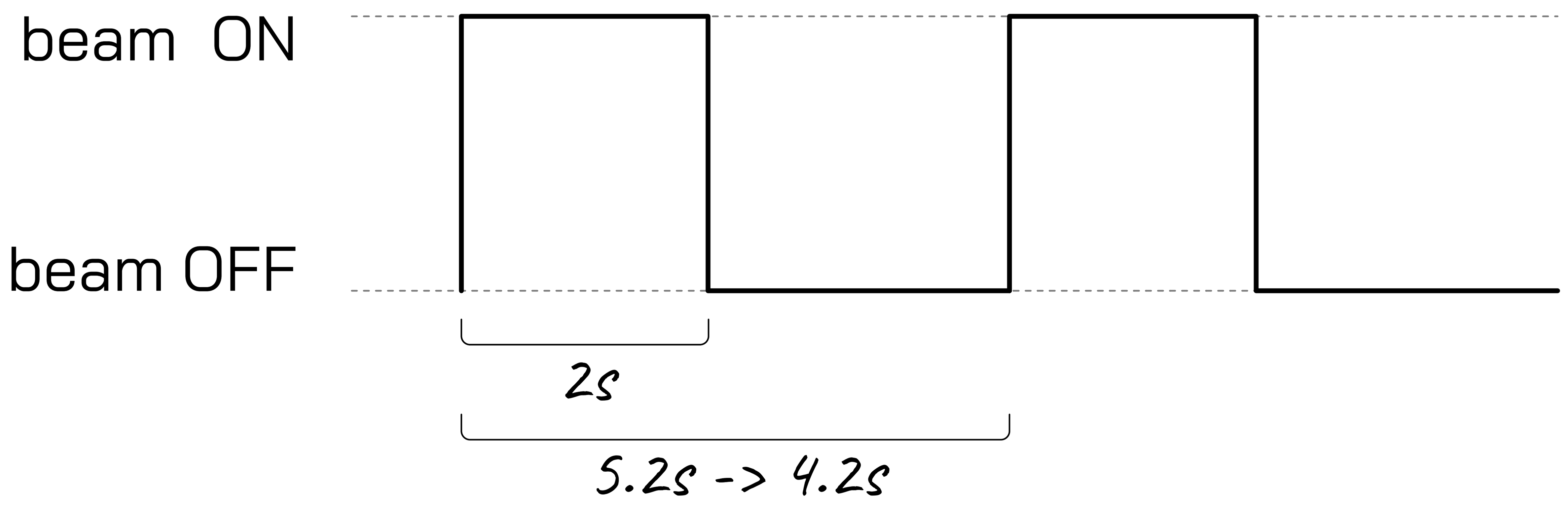
Analog waveforms





Overview of KOTO's DAQ system

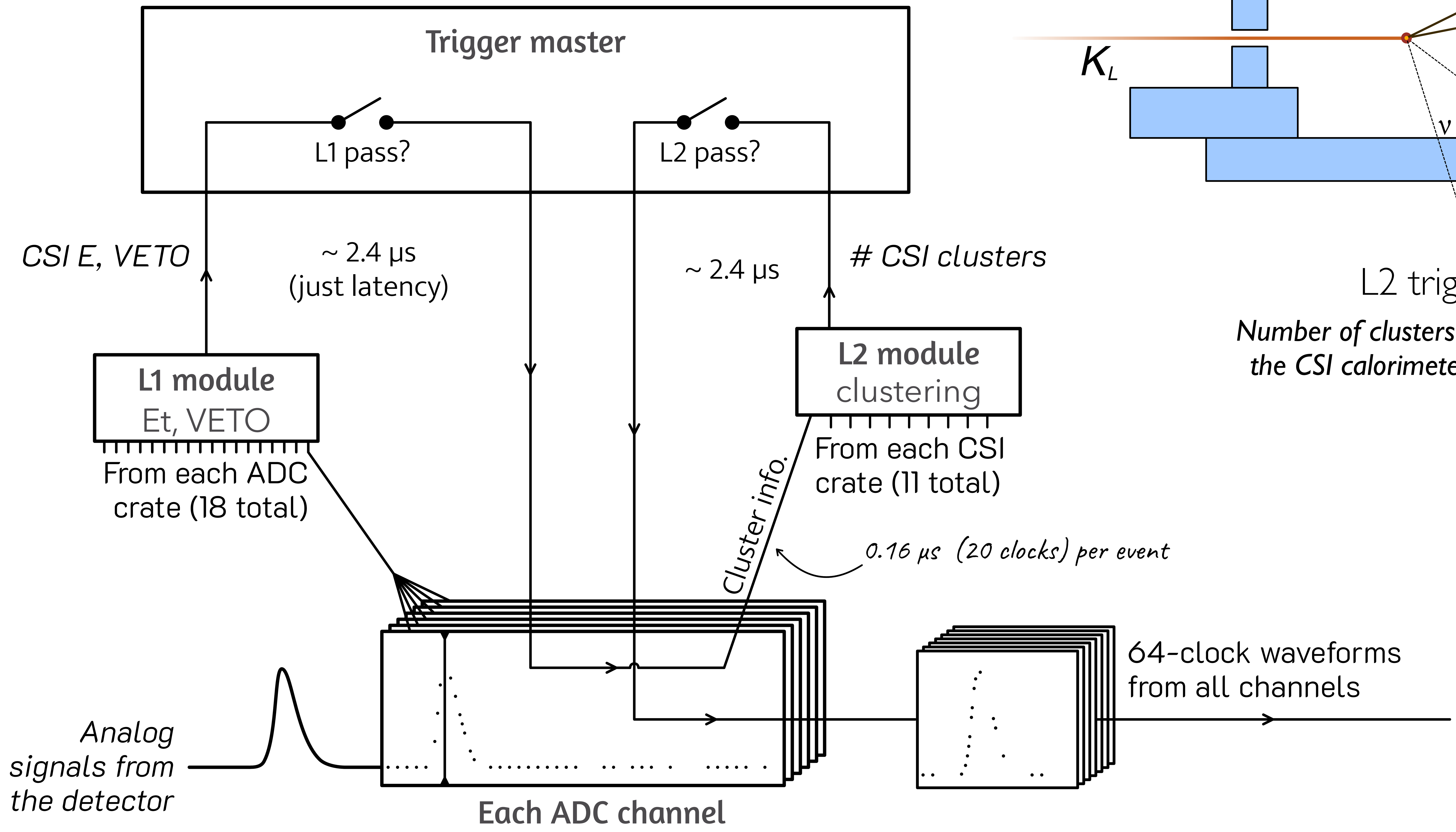
Spill cycle : Beam on-off cycle.

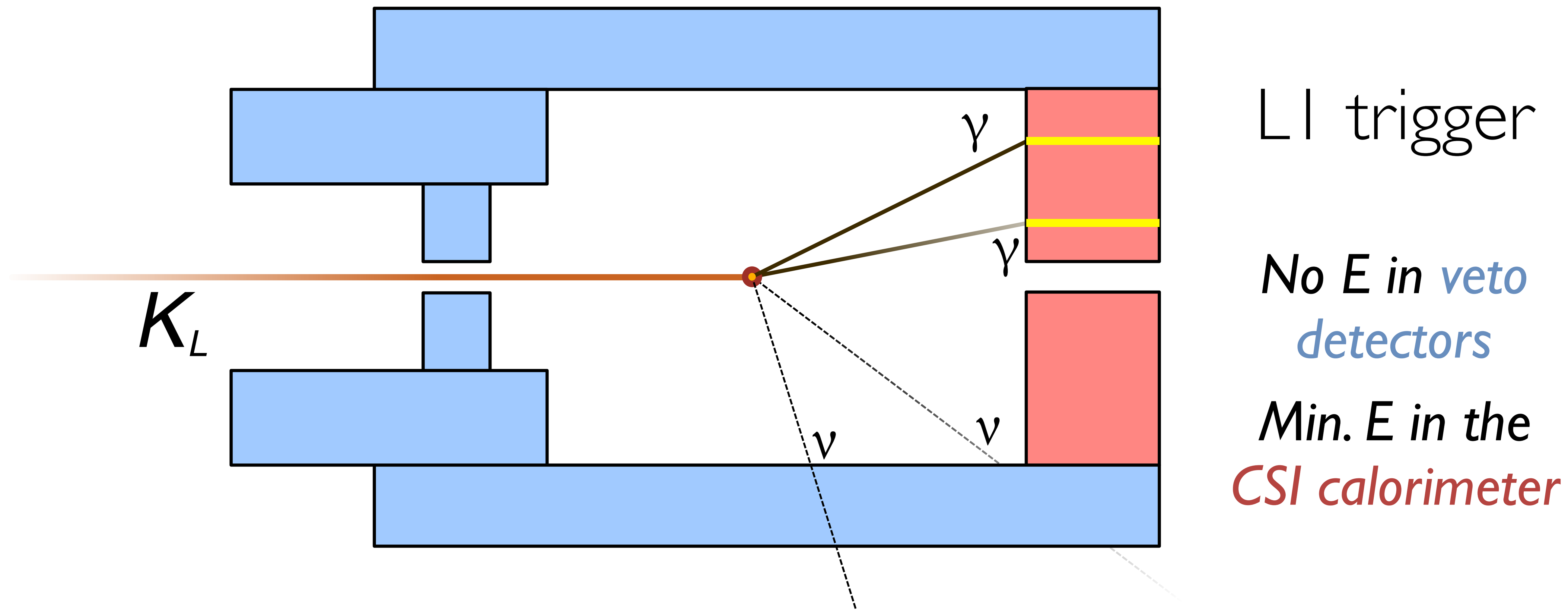


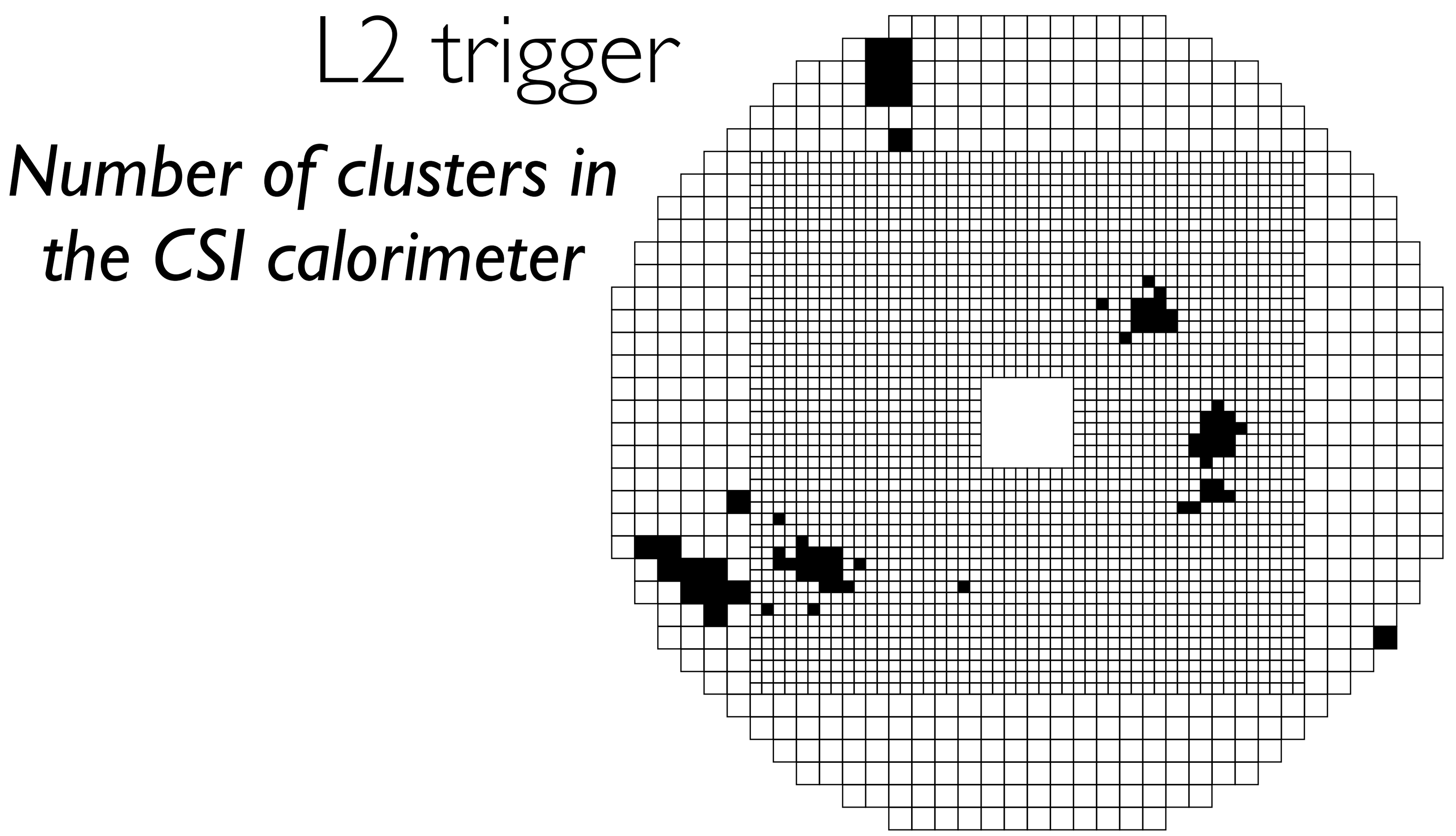
OFC: Optical Fiber center

Pipeline readout and trigger

Each ADC FPGA has enough memory to buffer data for 5.2 μs







16 channels per ADC module | 16 modules per ADC crate | 18 crates (11 just for the CSI calorimeter + 7 for VETO detectors)

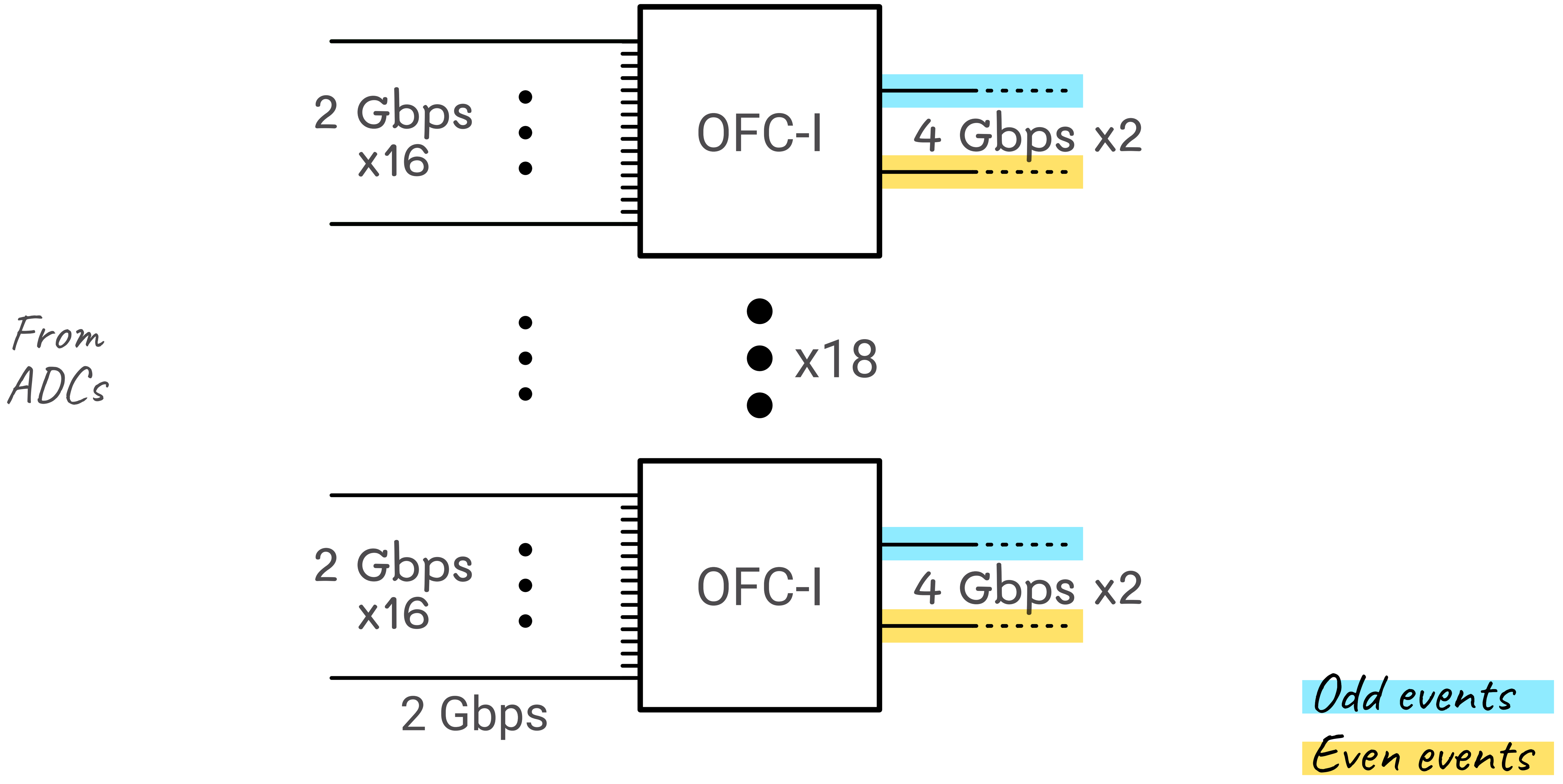
Event building and the OFC modules

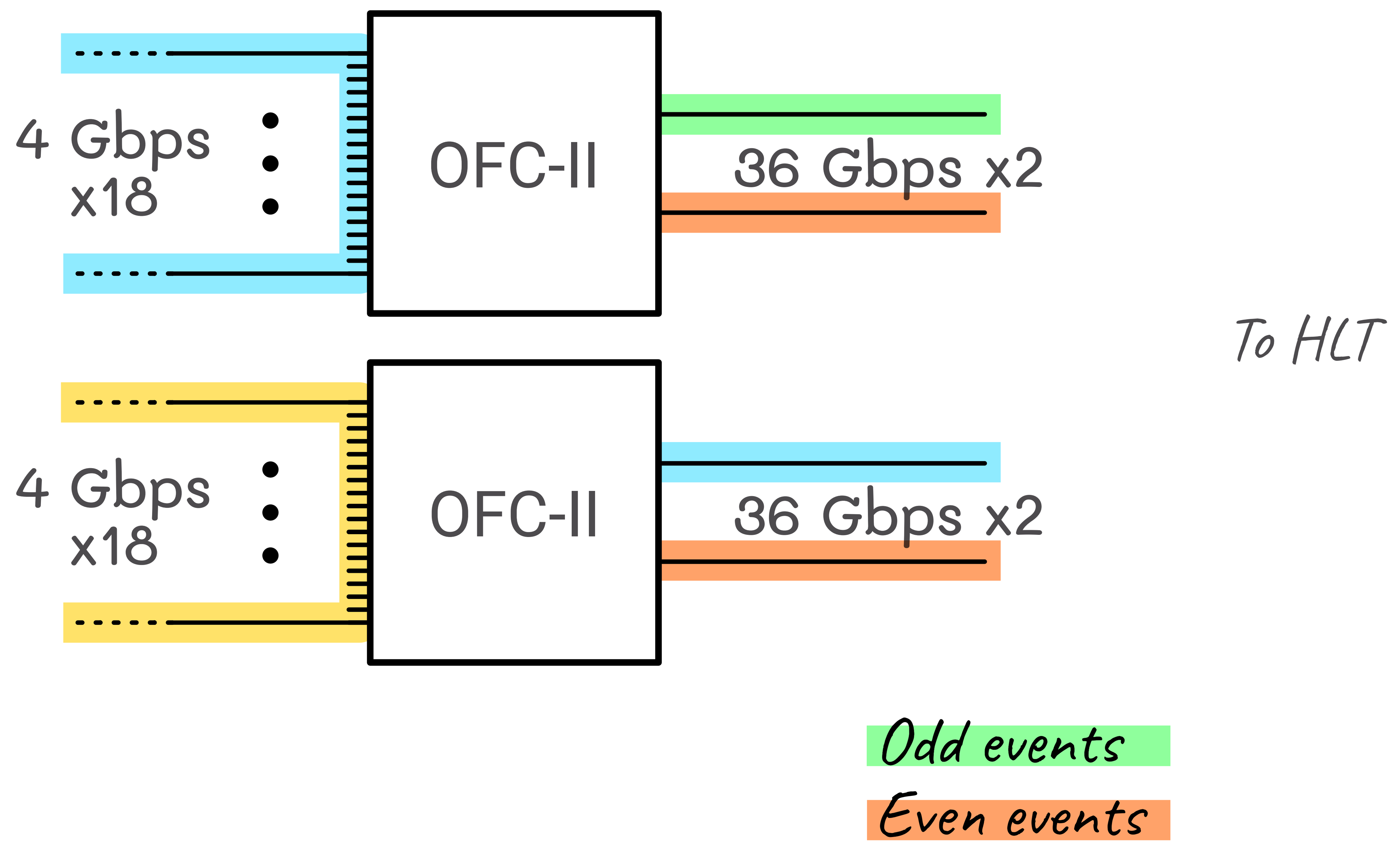
OFC-I

- Enough memory to hold 46 events
- **Up to 50 kEvents/spill with two OFC-IIs**
- Target OFC-II is switched event by event

OFC-II

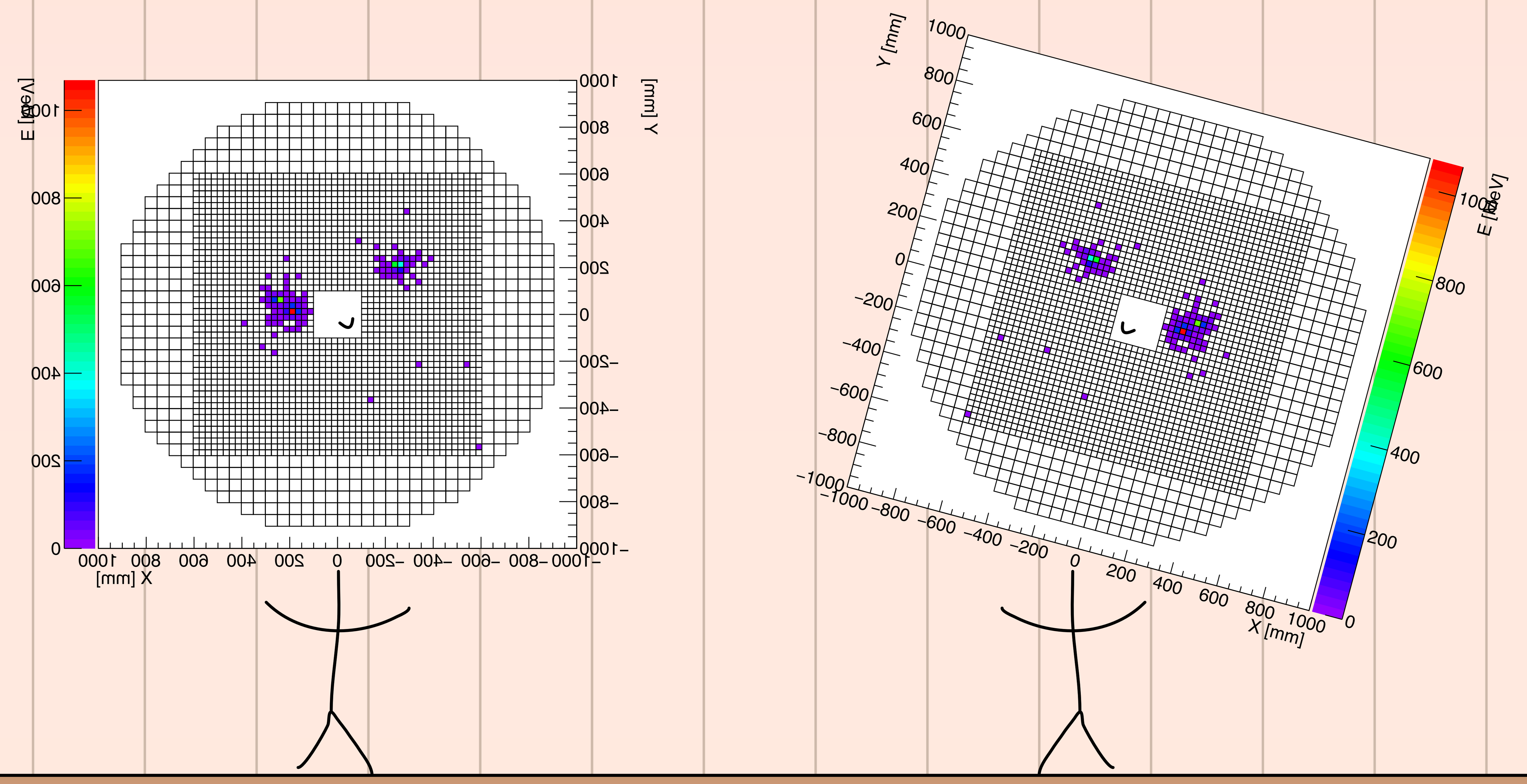
- Builds complete events from all OFC-I's data
- Sends them to the HLT through a 40 Gbps link
- Targets two HLT nodes per spill





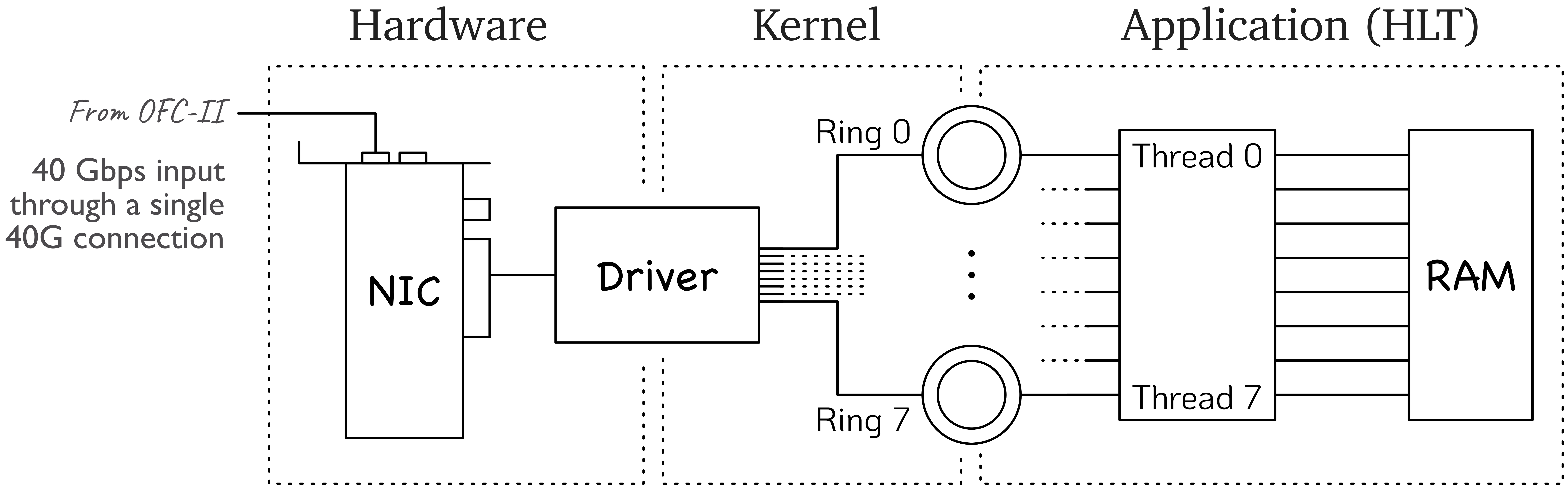
- The integrity of incoming data is checked every spill and every event.
- If checks don't pass or buffers get full, errors are issued and DAQ stops until the next spill.

The new GPU-based High Level Trigger of the KOTO experiment



40 Gbps data capture

Made possible thanks to Netmap, an open source framework for fast packet I/O[2]



8 threads take in parallel packets from the NIC into 8 ring buffers

8 HLT threads retrieve those packets and move them to much larger buffers on the RAM

The HLT nodes take advantage of the NUMA (Non Unified Memory Access) architecture:

- Threads involved in the 40G pcap are pinned to CPU cores with fastest access to the NIC
- HLT's RAM allocated in the memory region that those CPUs have fastest access to.

[2] https://github.com/luigirizzo/netmap

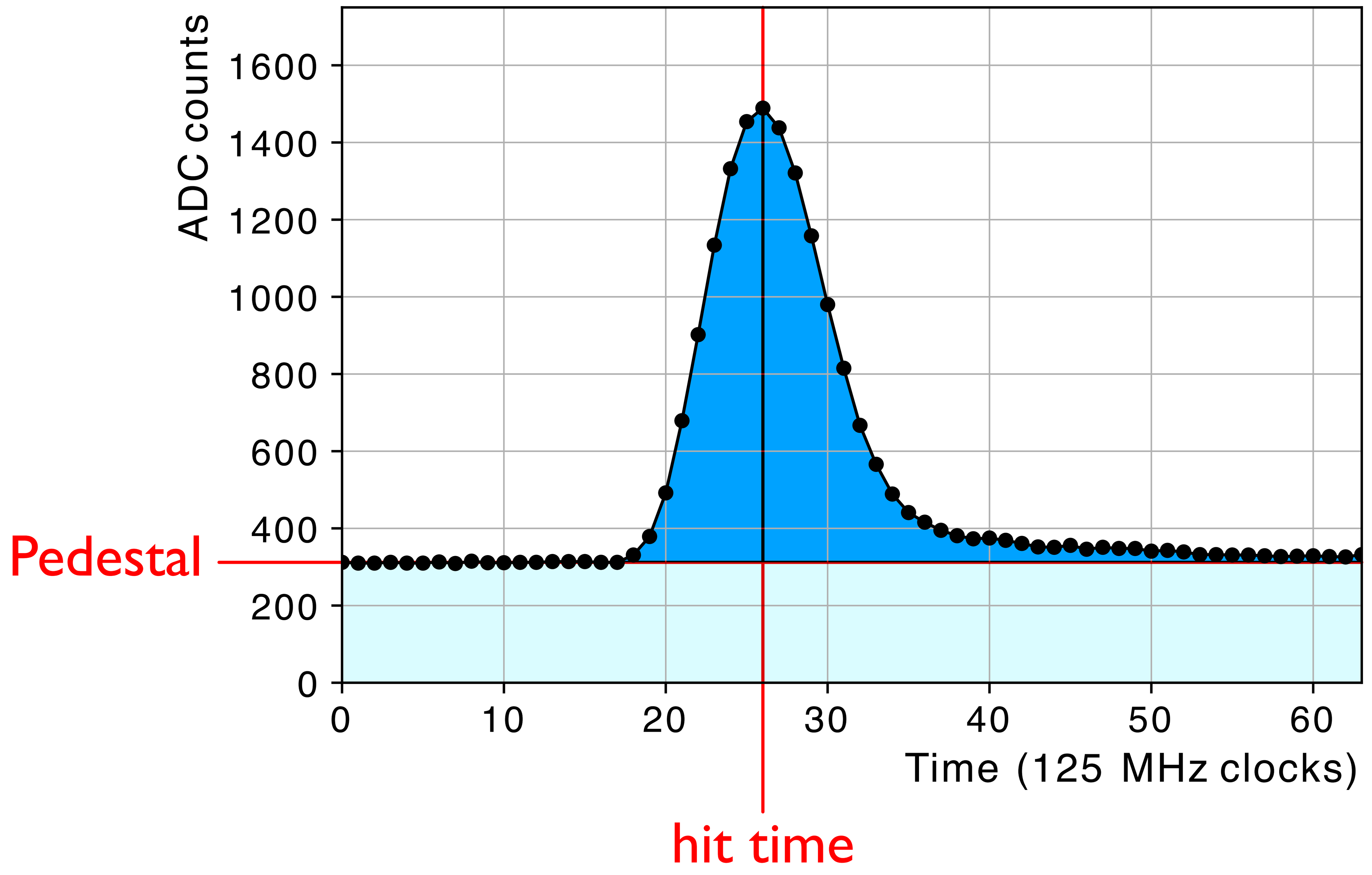
Event Reconstruction

- Based on CSI calorimeter data

Ch-by-Ch Energy

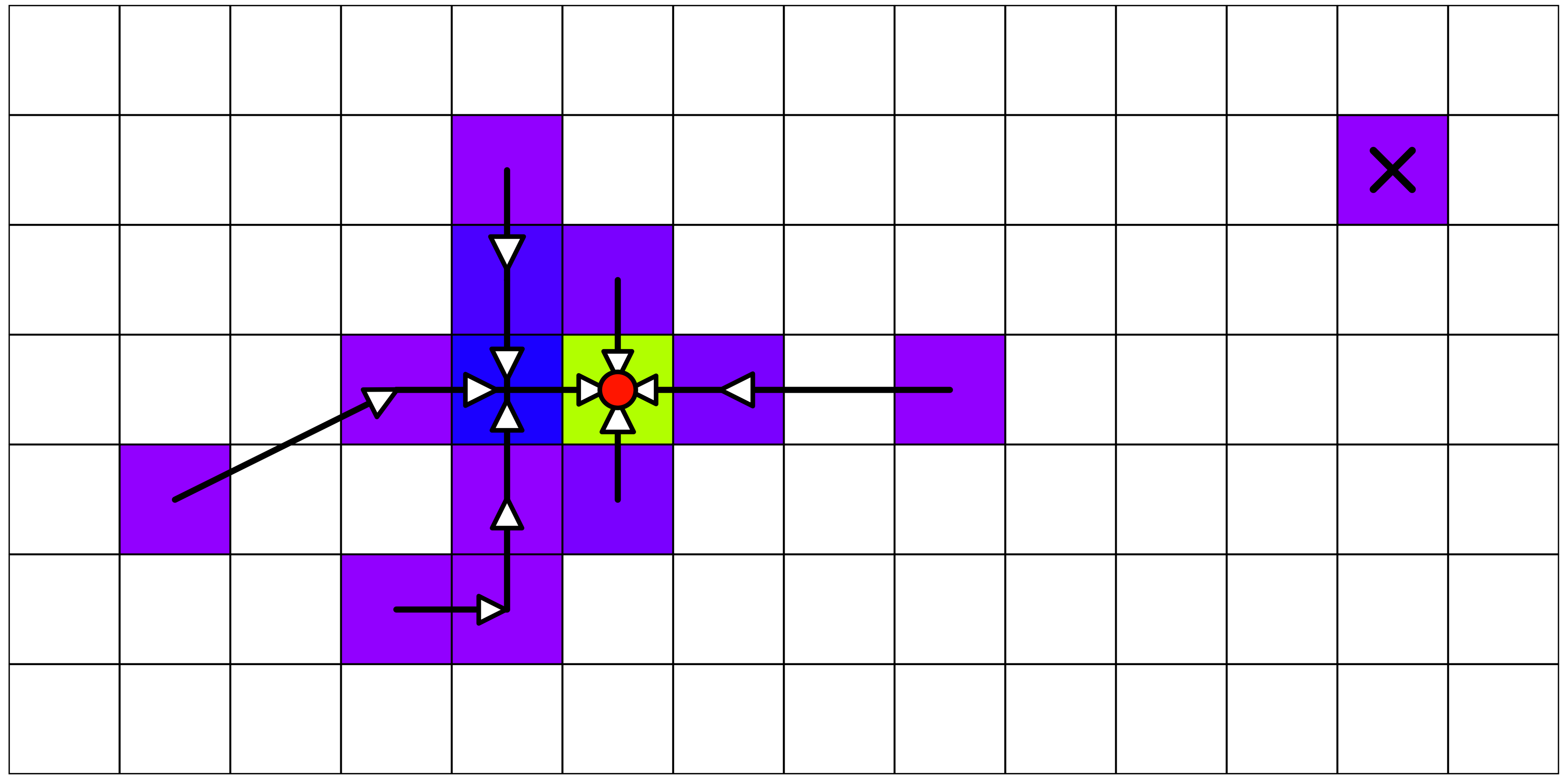
E = Integrated ADC x calibration constant

- *Energy calculated for all CSI calorimeter channels with on-time hits*
- *Calibration constants obtained from cosmic data before beam time*



Clustering

The clustering algorithm is an adaptation of the CLUE[1] algorithm, developed for CMSs new HGCAL:



1. Assign weights to crystals based on their energy (color)
2. Find the closest higher-E neighbor (arrows)

   Seeds (●): weight > threshold and no close neighbors with higher E

   Outliers (✕): weight < threshold and no close neighbors with higher E

3. Expand clusters from seeds

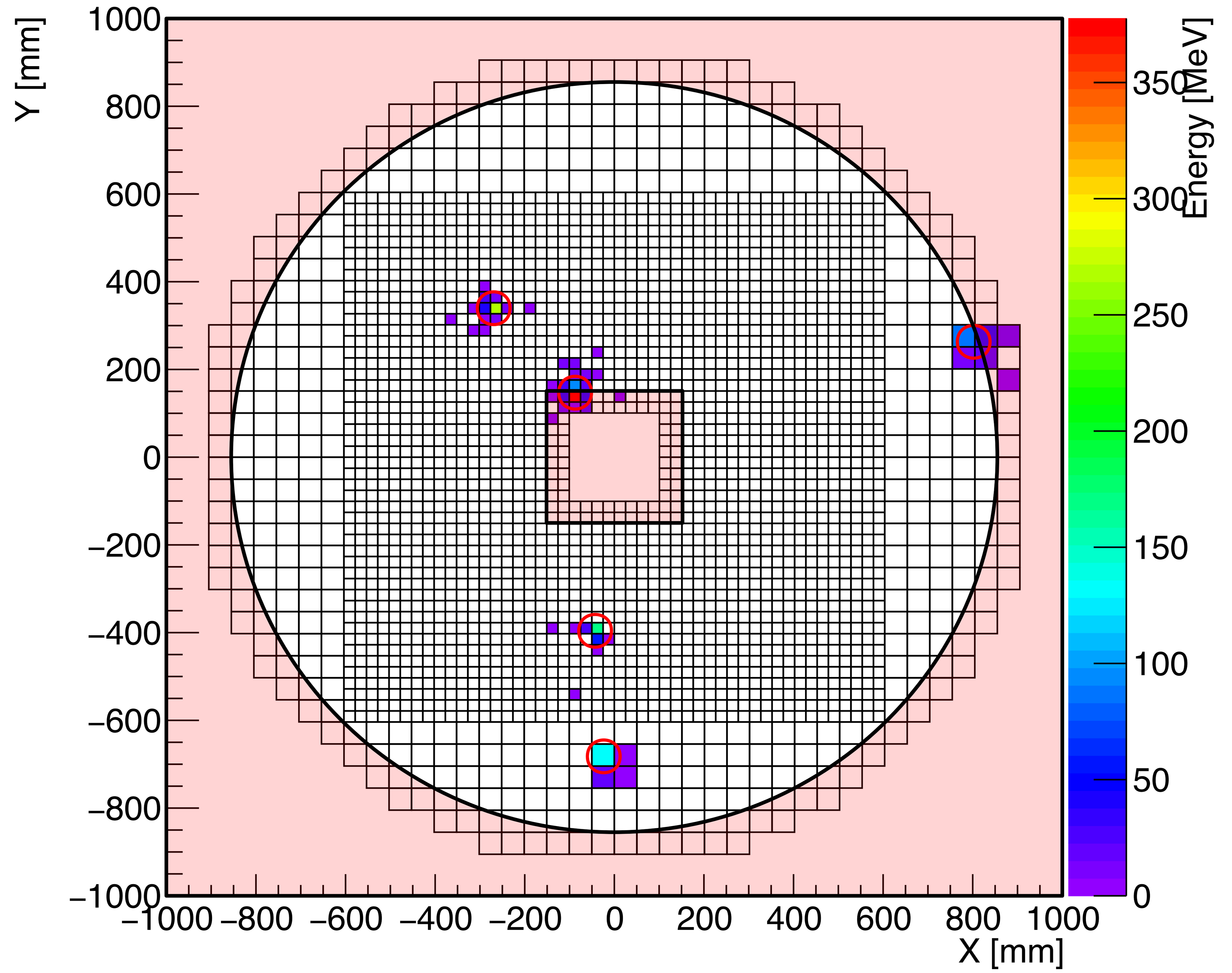
[1] doi: 10.3389/fdata.2020.591315

Event Selection

- Applied only needed loose cuts.

  No clusters in the shaded area

  Minimum total deposited energy in the calorimeter



- $$\text{L3 efficiency} = \frac{\text{events that pass the online and offline selection}}{\text{events that pass the offline selection}}$$

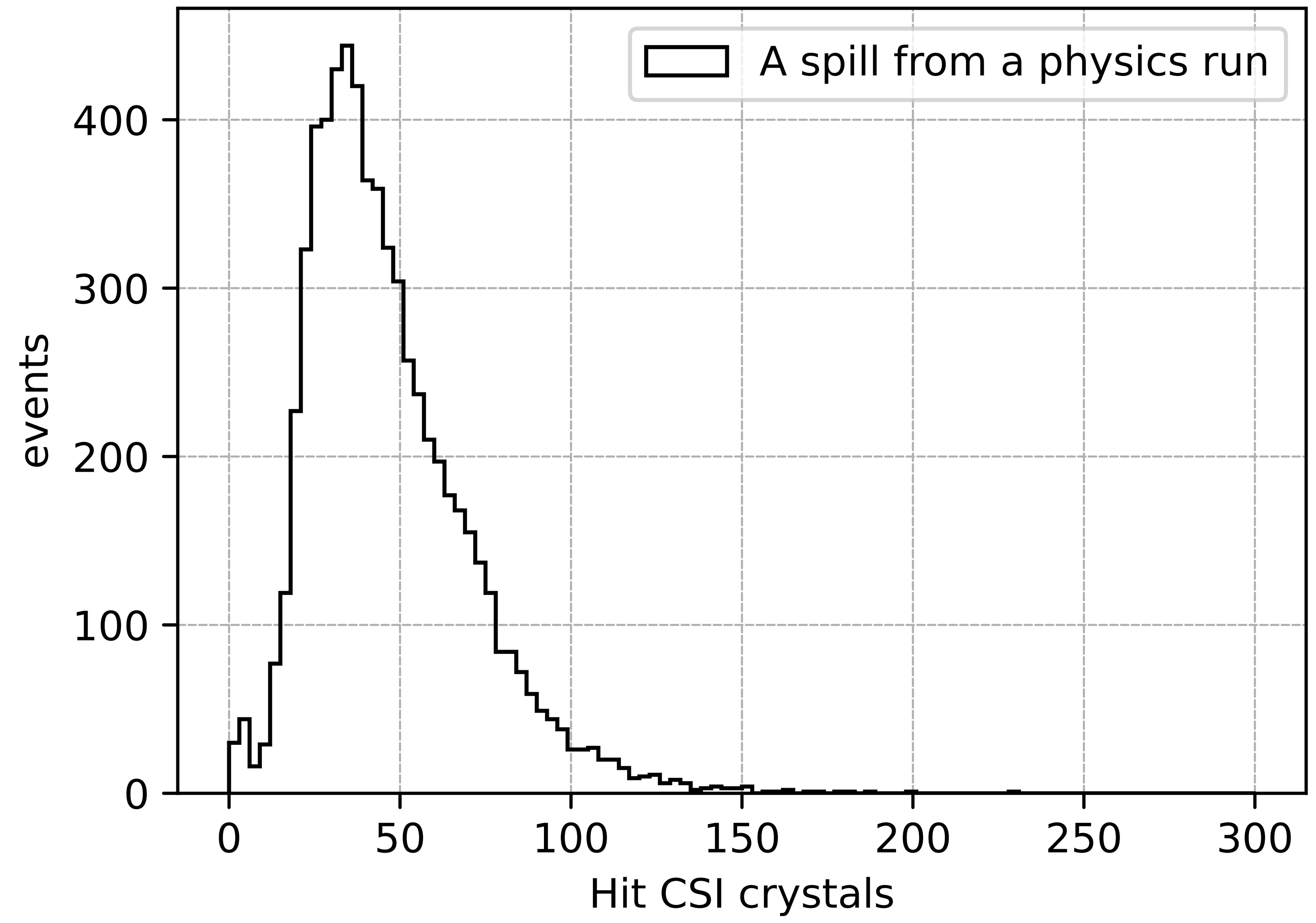
Trigger	HLT-input rate (Spring 2024)	Rate after event selection
$K_L \to \pi^0 \nu \overline{\nu}$	1.5 k/spill	Unchanged
$K_L \to 3\pi^0$ (6 clus.)	2.0 k/spill	Unchanged
$K^+ \to \pi^+ \pi^0$	5.7 k/spill	/ 1.25 (> 98.7% eff.)
$K_L \to 3\pi^0$ (5 clus.)	4.2 k/spill	/ 1.30 (> 98.7% eff.)
$K_L \to \pi^0 e^+ e^-$	2.4 k/spill	/ 1.20 (> 99.3% eff.)
Others	1.9 k/spill	Unchanged
Total	**17.7 k/spill (20.0 Gbps)**	**17.2 Gbps**

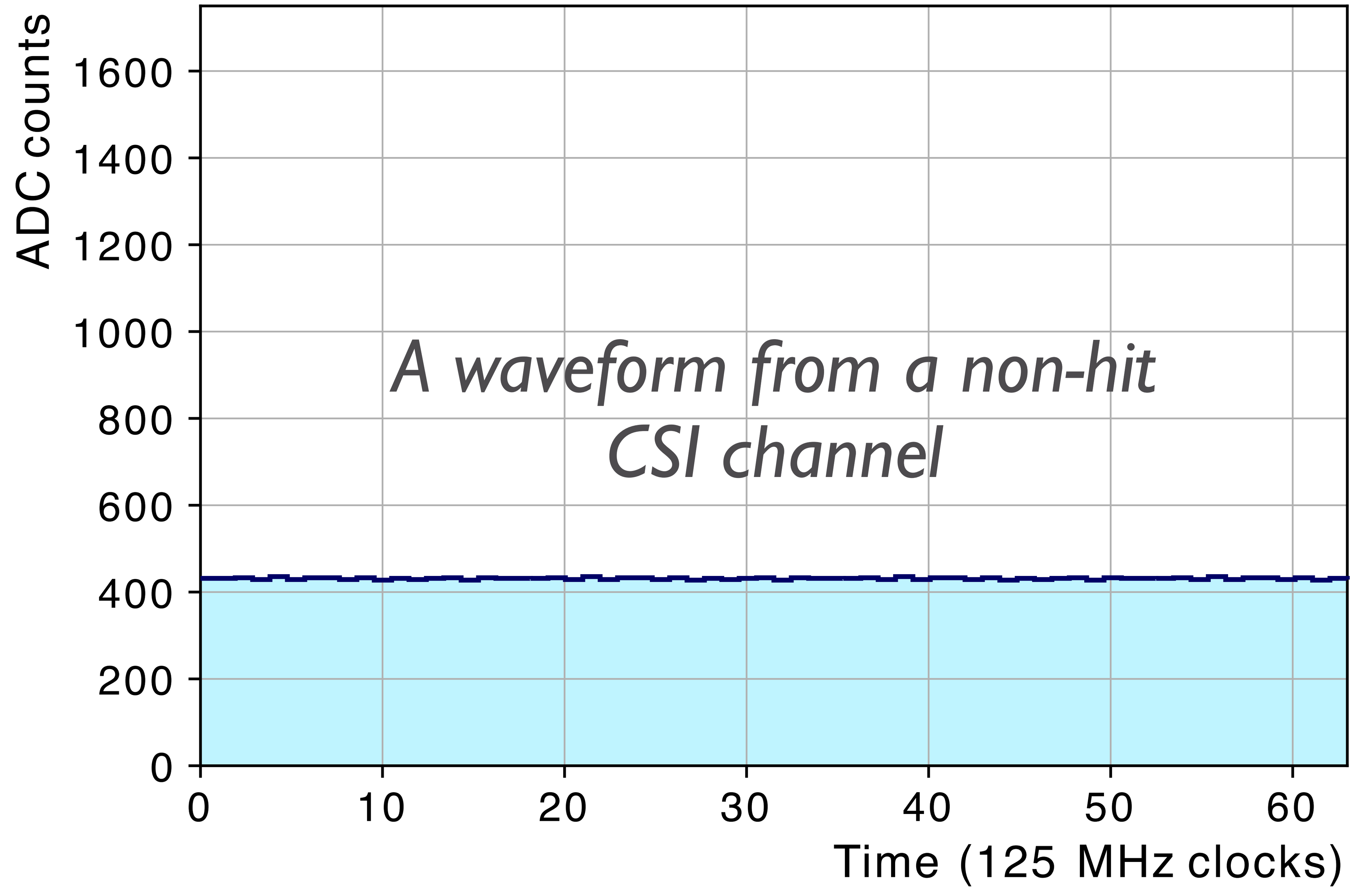
Tighter selection is not needed, thanks to the reduction coming from Pedestal Suppression and Waveform Compression

(Next two slides)

Pedestal Suppression

- Only ~40 of the almost 3000 CSI channels are hit per event
- Most channels without hit output very flat waveforms (noise) that do not contain relevant information.
- In practice, the suppression criteria is set to E ∈ (-2 MeV, 1 MeV)
- Exempt from being suppressed are:
    - Waveforms from the main physics trigger and other special triggers
    - Waveforms from all veto detectors.
    - Waveforms from low-gain CSI calorimeter channels

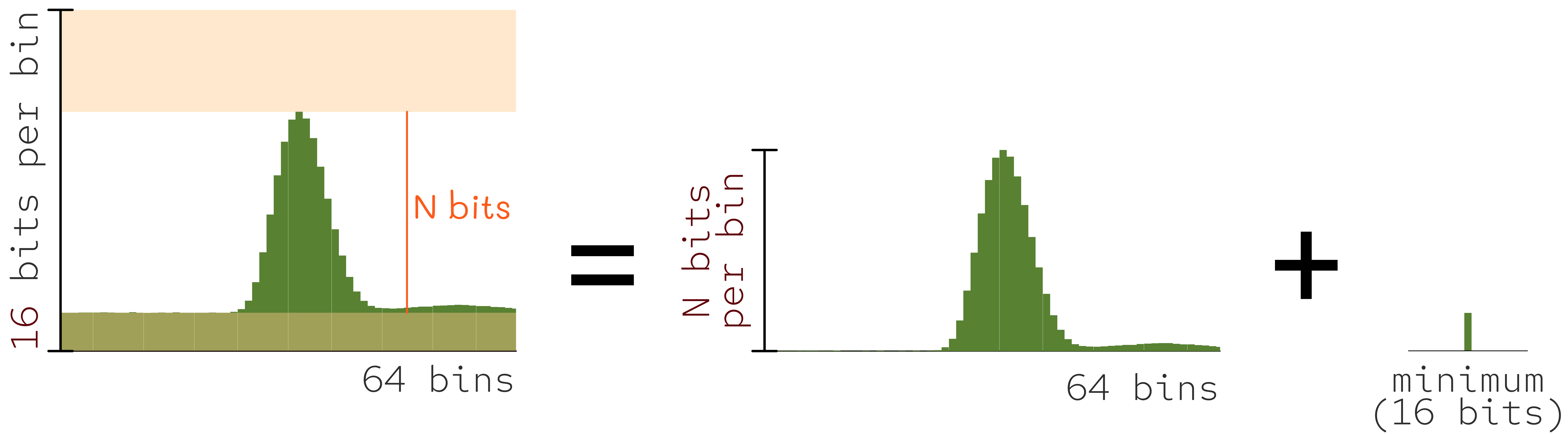




$$\text{P.S.inefficiency} = \frac{\text{waveforms accepted offline and suppressed online}}{\text{waveforms accepted offline}} \quad < 0.1\,\%$$

Waveform compression

- Conceptually very simple:



- Lossless
  - Applied to all waveforms of all events
- Powerful
  - Average compression factor of 3
- Very suitable for GPUs
  - No complex operations involved
  - Can be applied independently to all waveforms

Results: Data rate reduction at the HLT

Trigger	HLT-input rate (Spring 2024 physics runs)	Rate after event selection	rate after compression and ped. suppression
$K_L \to \pi^0 \nu \overline{\nu}$	1.5 k/spill	Unchanged	/ 3.2
$K_L \to 3\pi^0$ (6 clus.)	2.0 k/spill	Unchanged	/ 5.3
$K^+ \to \pi^+ \pi^0$	5.7 k/spill	/ 1.25	/ 5.4
$K_L \to 3\pi^0$ (5 clus.)	4.2 k/spill	/ 1.30	/ 5.3
$K_L \to \pi^0 e^+ e^-$	2.4 k/spill	/ 1.20	/ 5.3
Others	1.9 k/spill	Unchanged	/ 4.1
Total	**17.7 k/spill (20.0 Gbps)**	**17.2 Gbps**	**3.6 Gbps**

Largest reduction factors come from pedestal suppression and waveform compression

No strict selection needed this time to overcome the J-PARC to KEK bandwidth bottleneck

Conclusion

KOTO has successfully taken data this Spring after a major DAQ upgrade

- ▹ Event building performed in FW, so the HLT gets complete events.
- ▹ Event selection and further data reduction implemented on GPU at the HLT

Together with the main $K_L \to \pi^0 \nu \bar{\nu}$ data, KOTO is able to collect for the first time $K_L \to 3\pi^0$ (5 hits on the calorimeter) to study veto inefficiencies, and $K_L \to \pi^0 e^+ e^-$, to study the feasibility of its future BR measurement.

The current DAQ HW has the potential to take physics data at up to 50 kEvents / spill

- ▹ x2.5 higher than current rate

Large margin to tighten the current event selection and/or to add more cuts at the HLT

Backup

Error monitoring at the OFC modules

Optical Link error:

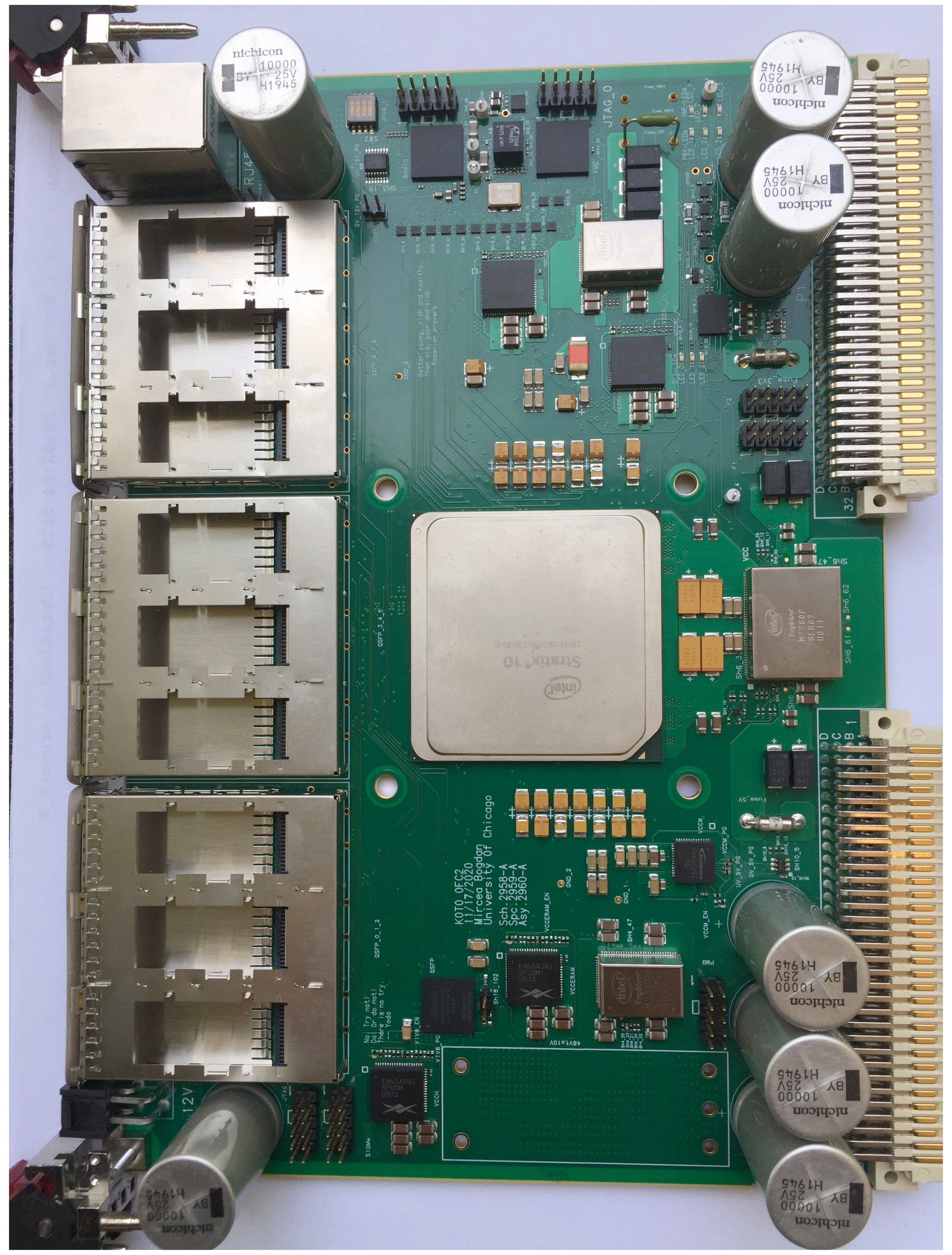
Known data is received and checked at the beginning of every spill

Data alignment error:

whether data has been received from all inputs is checked event by event

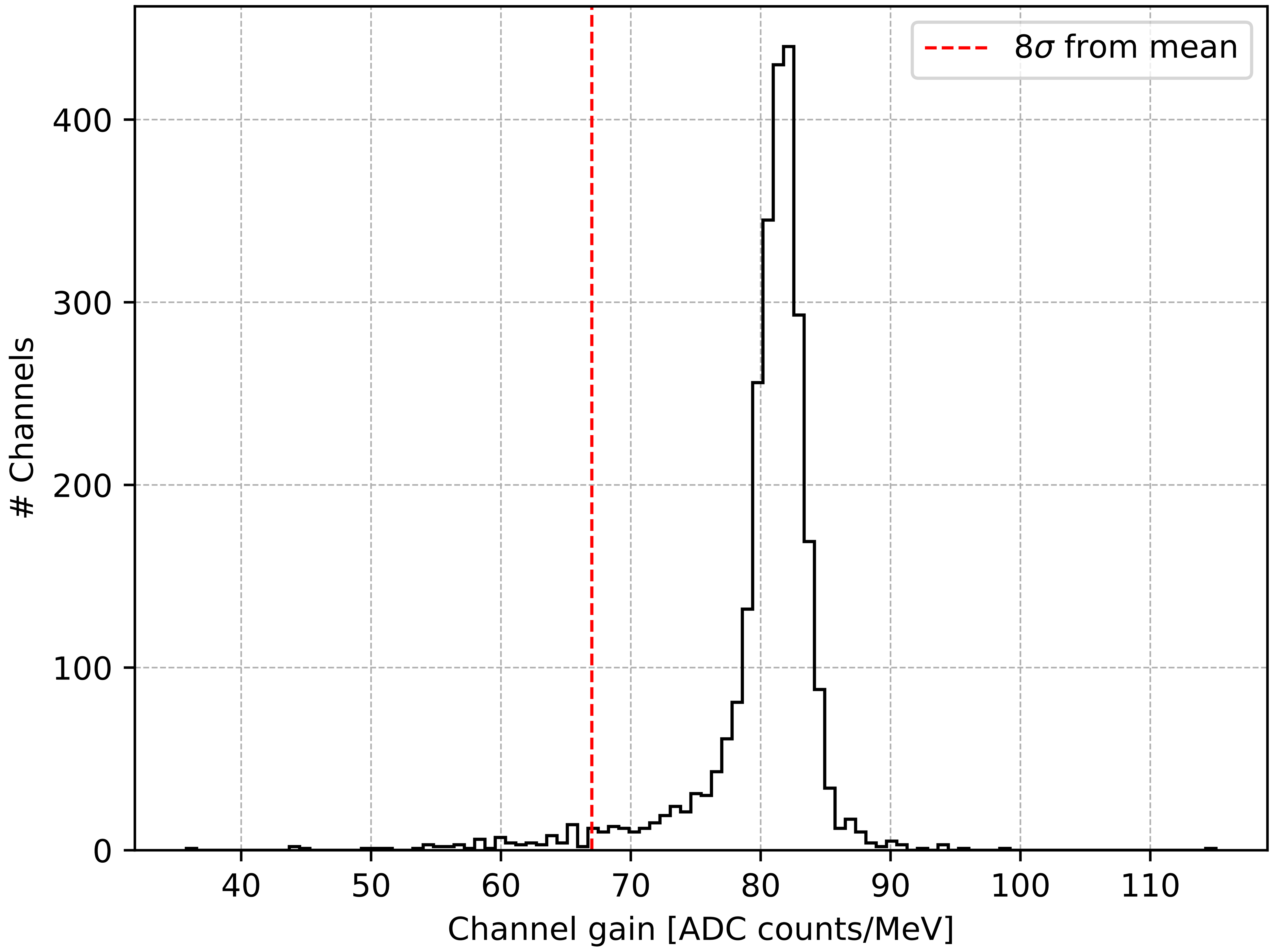
Busy error:

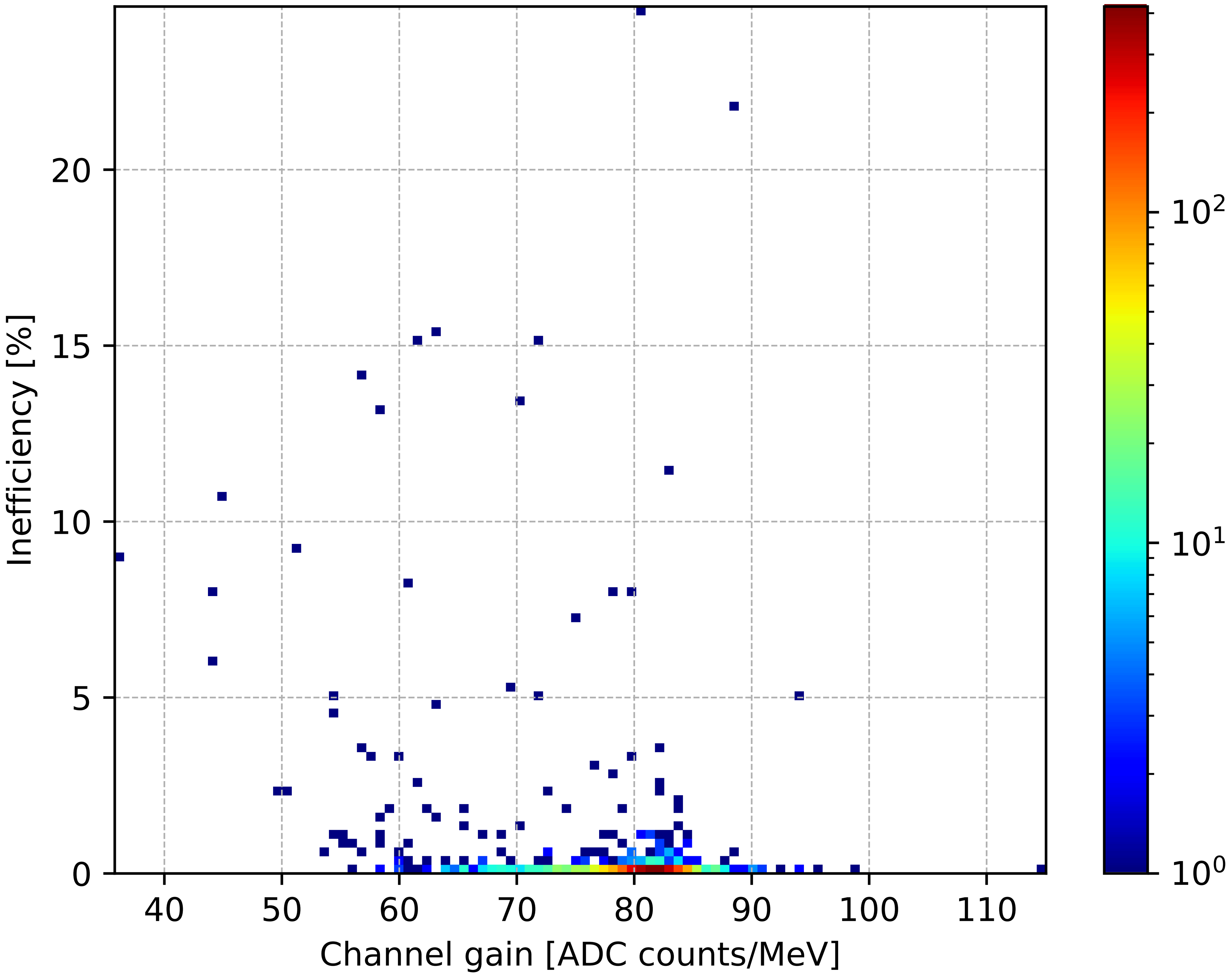
Issued when input > output and memory starts becomes full



Pedestal suppression inefficiency

Channels with low gain are masked, as their low peak/noise ratio makes the PS less efficient

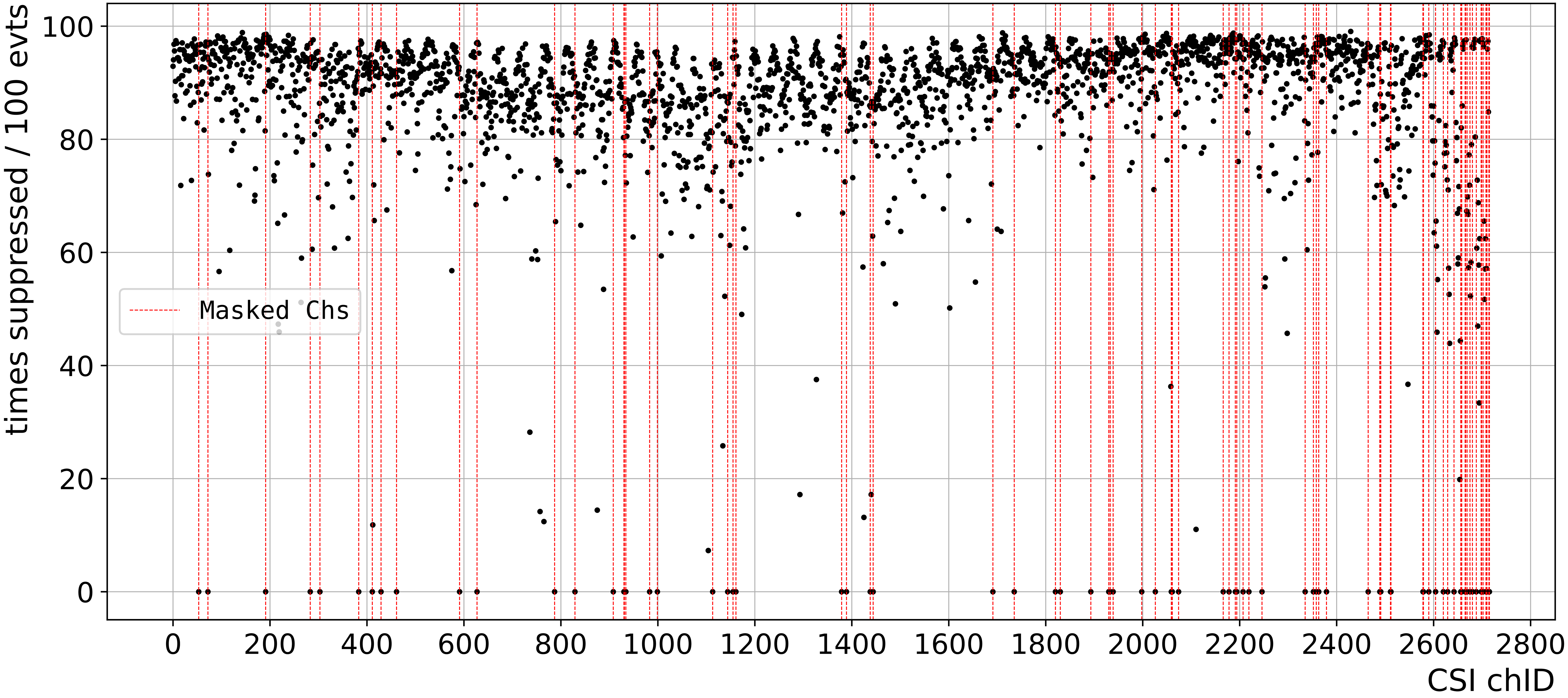




Pedestal suppression inefficiency: **Results in physics runs**

Average channel PS inefficiency after masking low-gain channels is 0.09 %

Pedestal suppression inefficiency: **Results in physics runs**



Average suppression rate is 86 %

When the E criteria fails

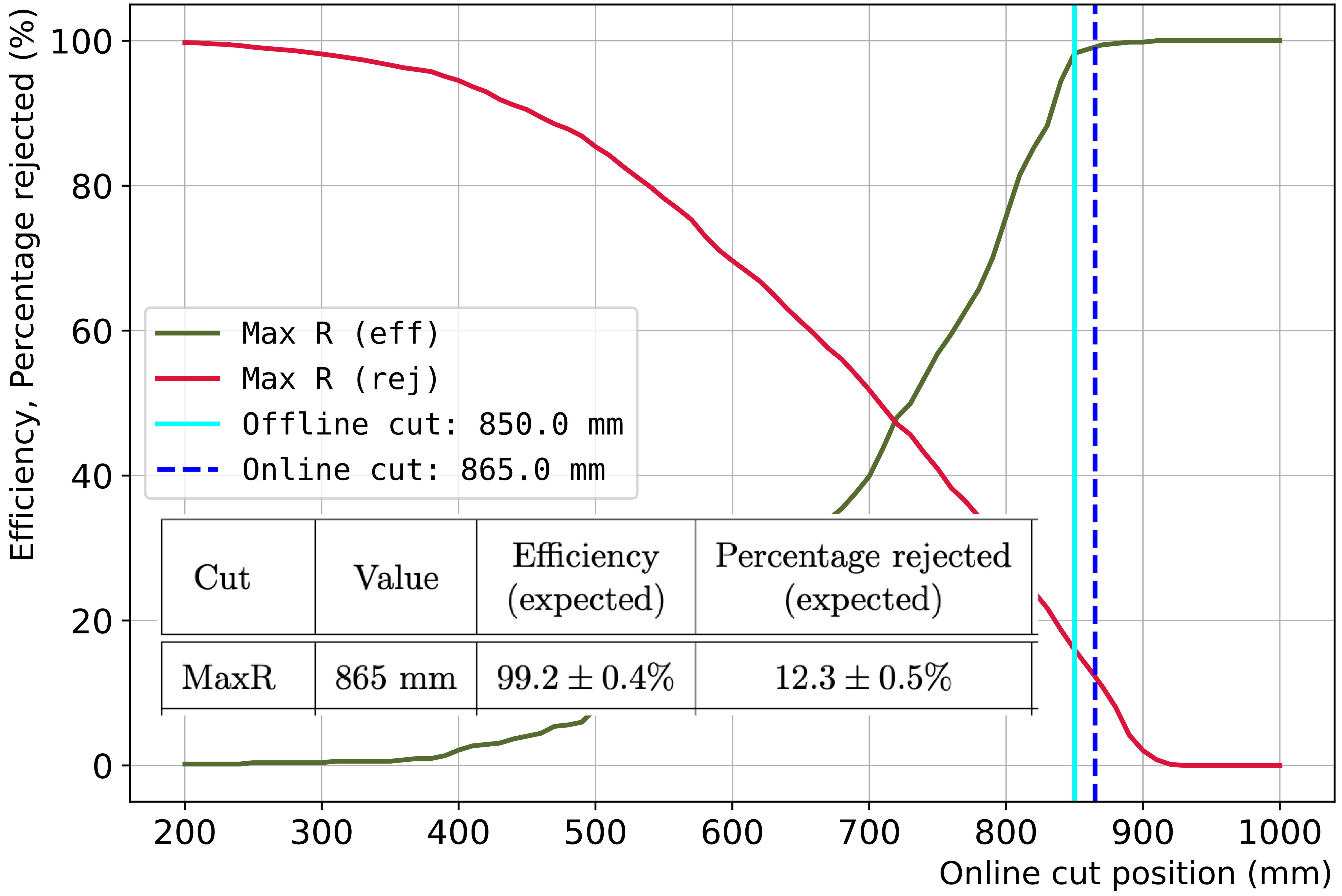
PH < 10 counts (roughly 1 MeV) imposed together with the E criteria

Event reconstruction and event selection efficiencies

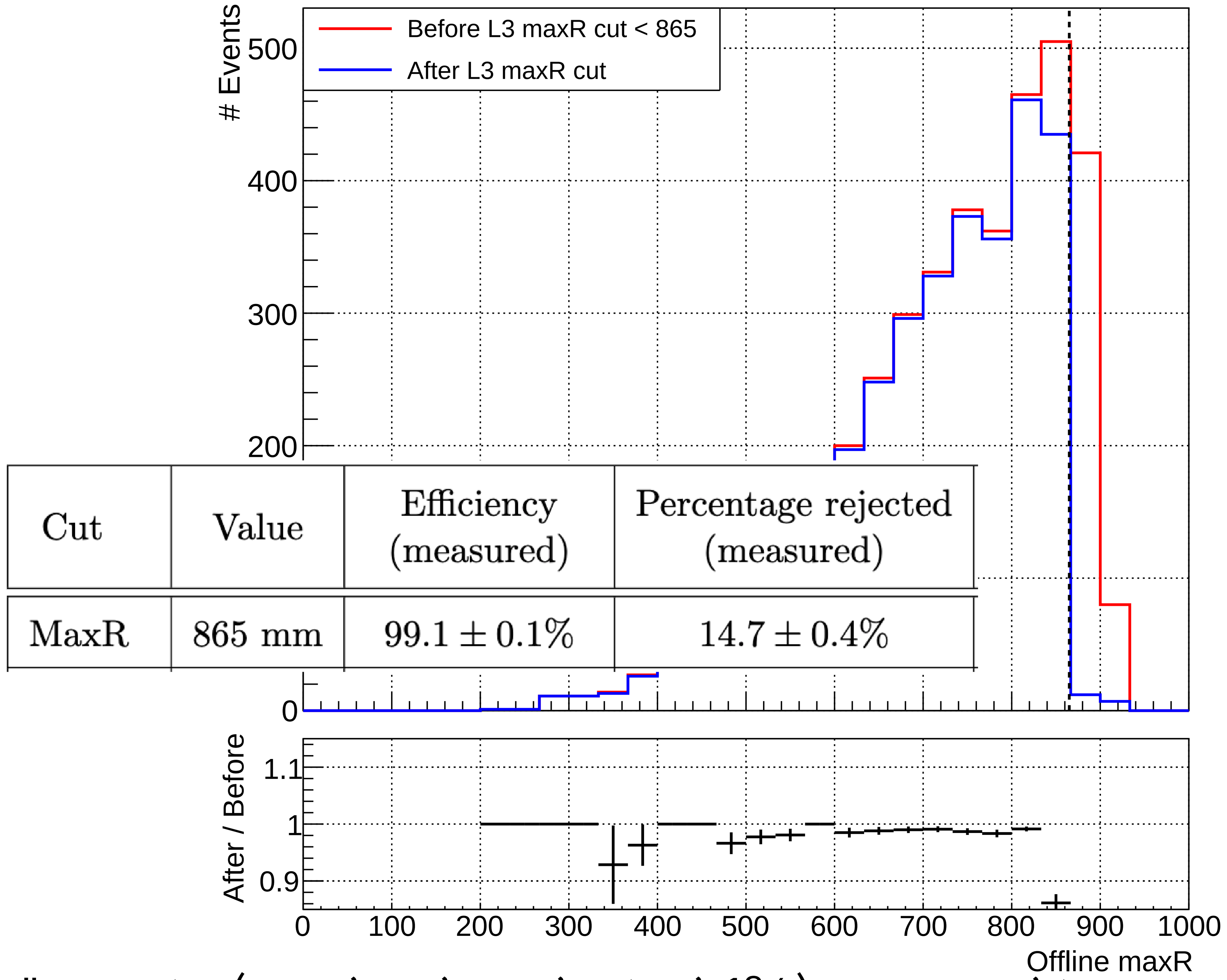
$$\text{L3 efficiency} = \frac{\#\ \text{events that pass the online and offline selection}}{\#\ \text{events that pass the offline selection}}$$

1: Non-selected data -> to OFC2 format -> Fed back to the L3 -> calculate eff. for different thresholds. (left fig.)

2: Thresholds are put into the L3 sw -> special "tagging" runs are taken -> expectations are verified during actual physics runs (right fig.)



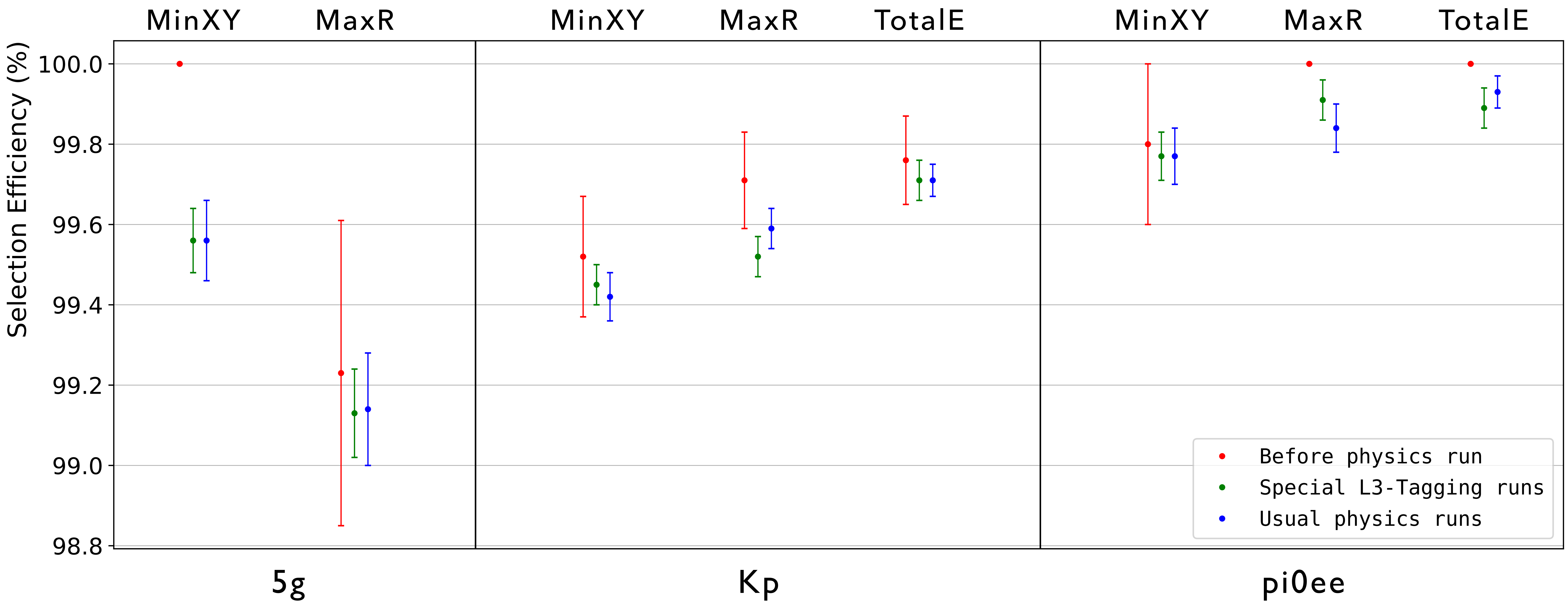
Cut	Value	Efficiency (expected)	Percentage rejected (expected)
MaxR	865 mm	$99.2 \pm 0.4\%$	$12.3 \pm 0.5\%$



Cut	Value	Efficiency (measured)	Percentage rejected (measured)
MaxR	865 mm	$99.1 \pm 0.1\%$	$14.7 \pm 0.4\%$

3: L3 thresholds fixed -> L3-selection is enabled. "L3-min-bias" events (randomly selected 1%) are used to re-measure the inefficiencies during real physics runs

Event reconstruction and event selection efficiencies



Trigger	Efficiency	Rejected fraction
5γ	$98.7 \pm 0.2\%$	$23.0 \pm 0.3\%$
K^+	$98.7 \pm 0.1\%$	$20.0 \pm 0.2\%$
$K_L \to \pi^0 e^+ e^-$	$99.3 \pm 0.1\%$	$16.0 \pm 0.3\%$

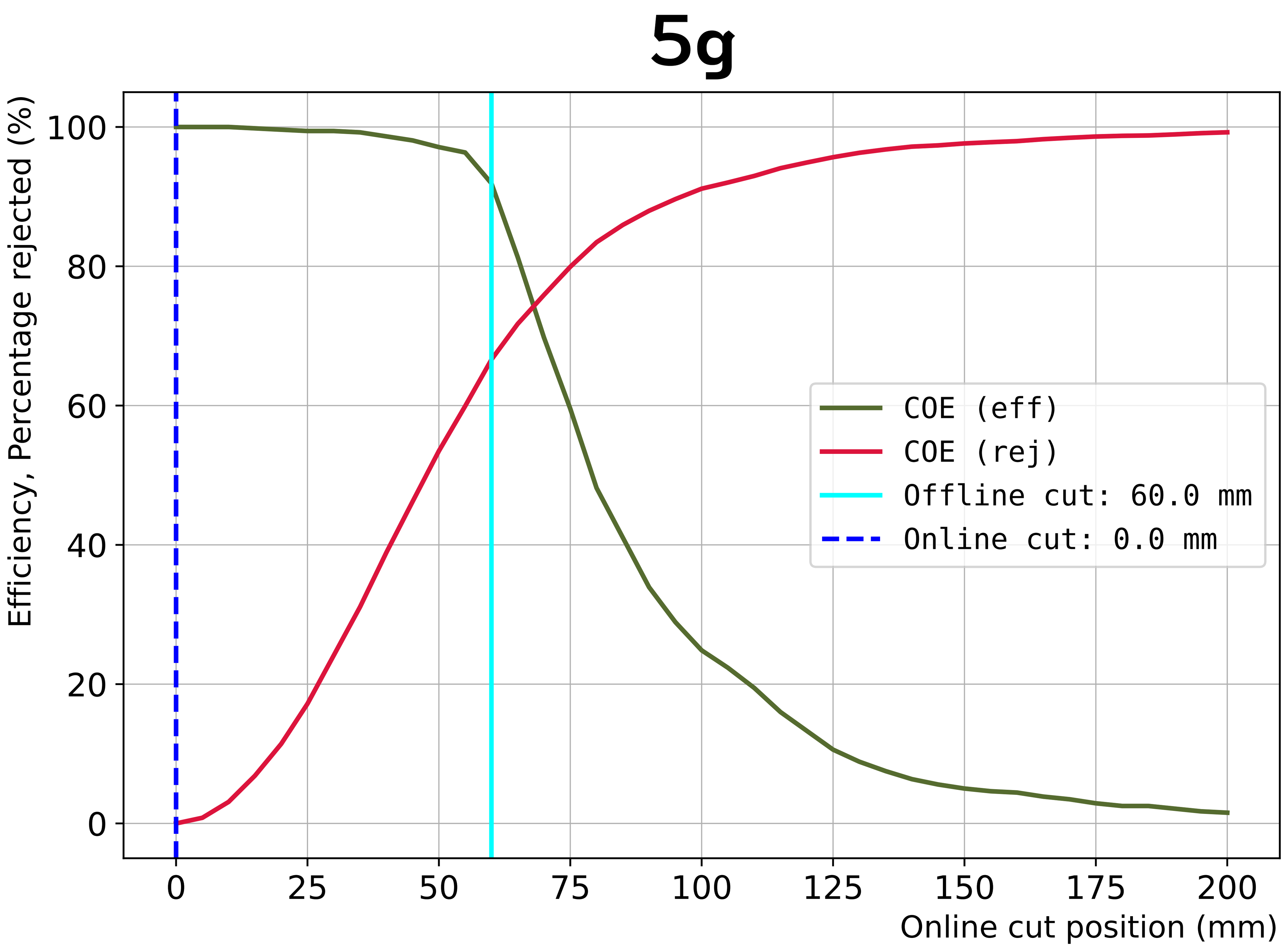
Table 9.1: Combined efficiency and rejection power of all L3 cuts applied to the 5γ, K^+ and $K_L \to \pi^0 e^+ e^-$ triggers.

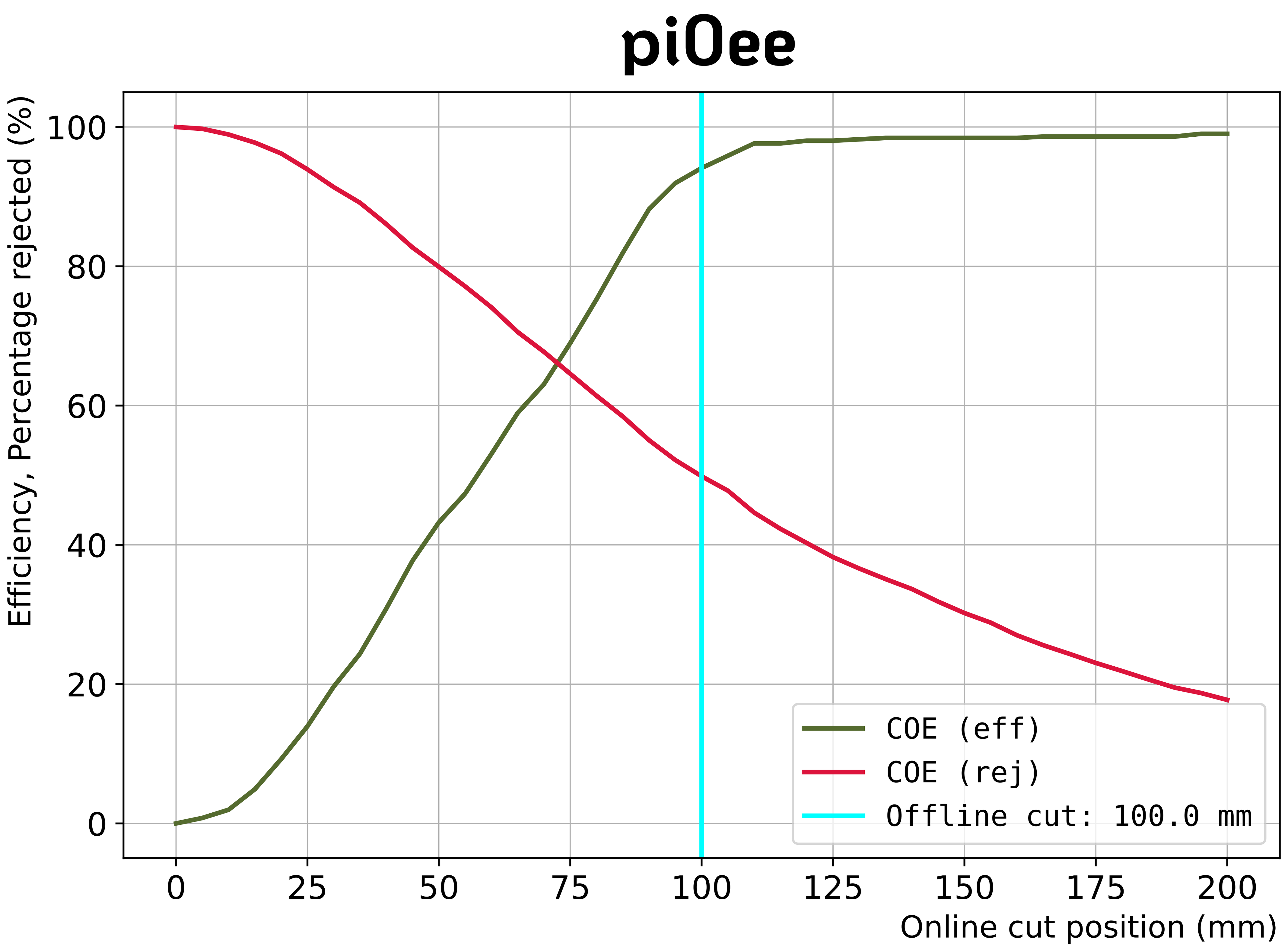
Efficiencies were high and stable during the run. (green is right after implementing L3 selection, blue is close to the end of beamtime)

Notes about online L3 reconstruction and selection

- Only fiducial (MinXY and MaxR) cuts applied to 5g, Fid + TotalE cuts are applied to K+ and pi0ee

- COE cuts for pi0ee and 5g were not applied, as the data rate was found to be within requirements even without the cuts. They still have great potential and could be used in the future.





- The denominator of the efficiency does *not* include offline cluster-shape and mode-specific offline cuts.
- The efficiency calculation implies that the offline reconstruction is perfect.

DAQ efficiency in run91

L3 efficiency

= *nEvents recorded offline / nEvents at OFC-II output*, using only spills where L1A and OFC-II output are equal.

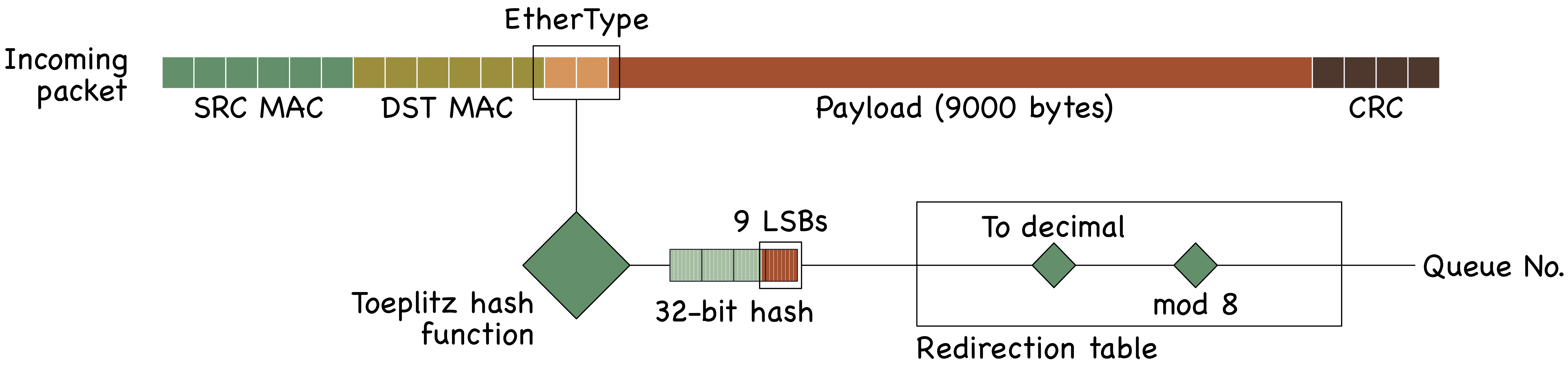
DAQ efficiency in run91

Overall DAQ efficiency

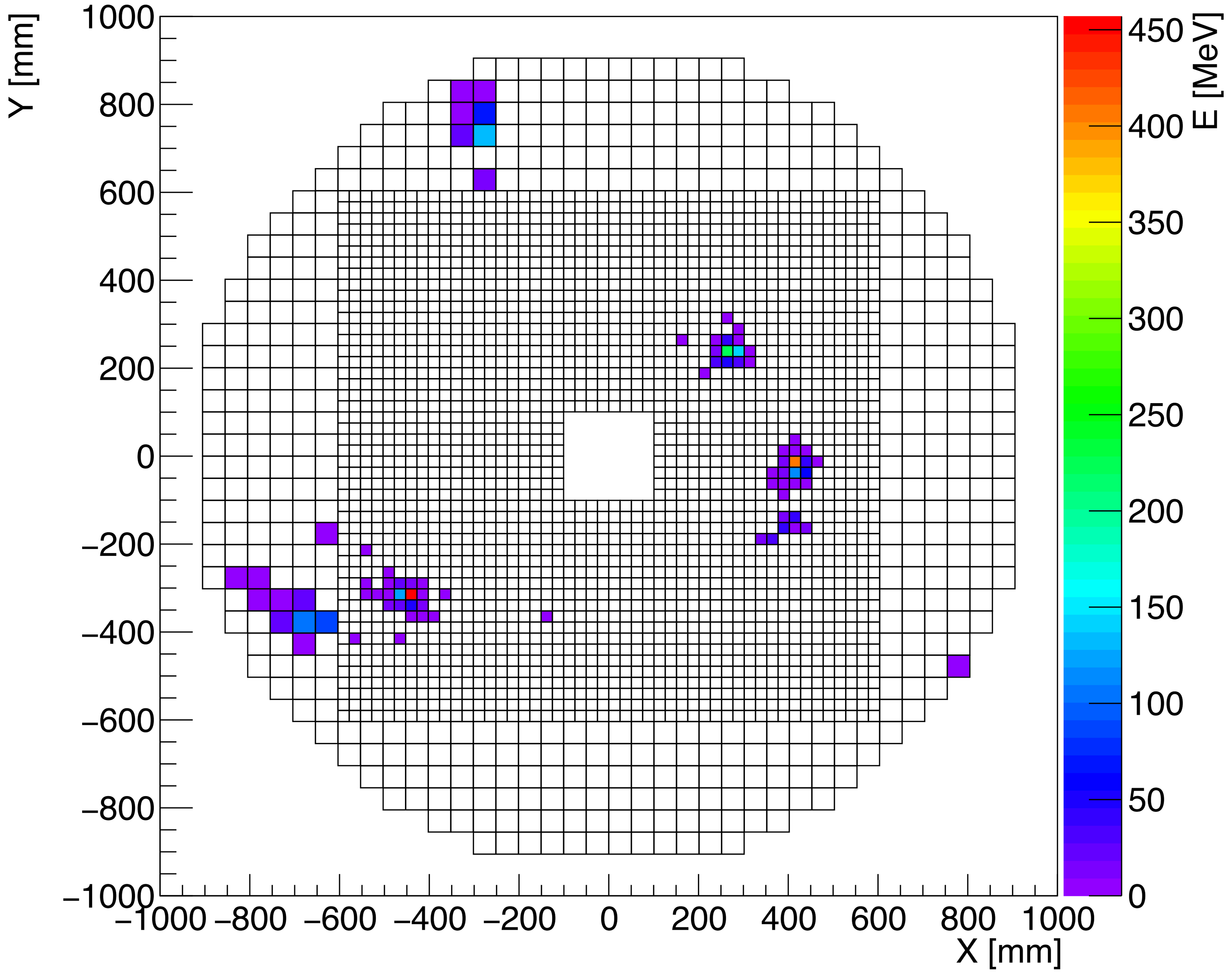
RDMA

CPU
PCAP RAM
COMPUTE RAM
SPILL NODES
DISK NODES
CPU
RAM
40G
PCIe SWITCH
10 G
10 G
PCIe SWITCH
VRAM
GPU
Disks

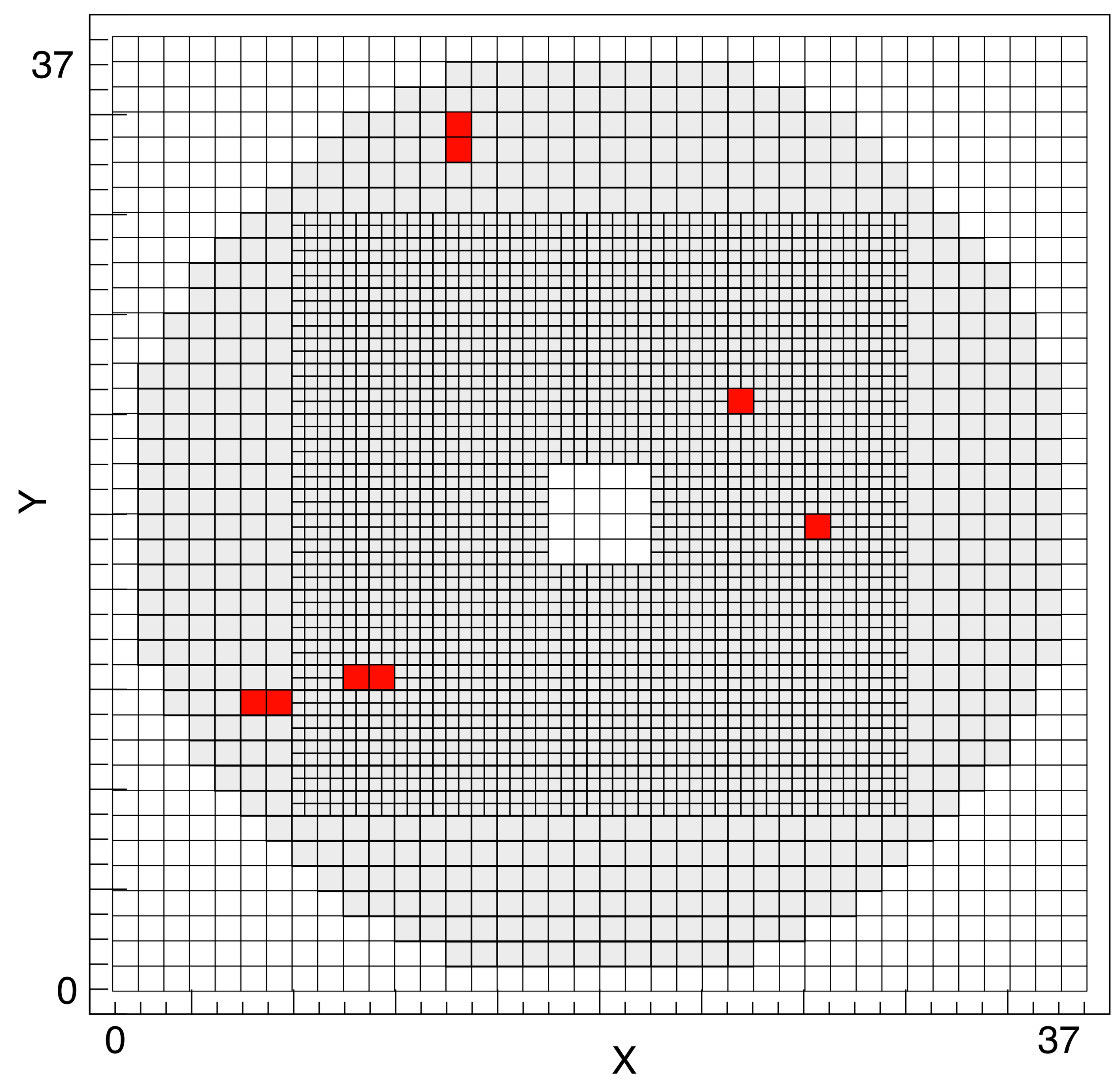
packet redirection

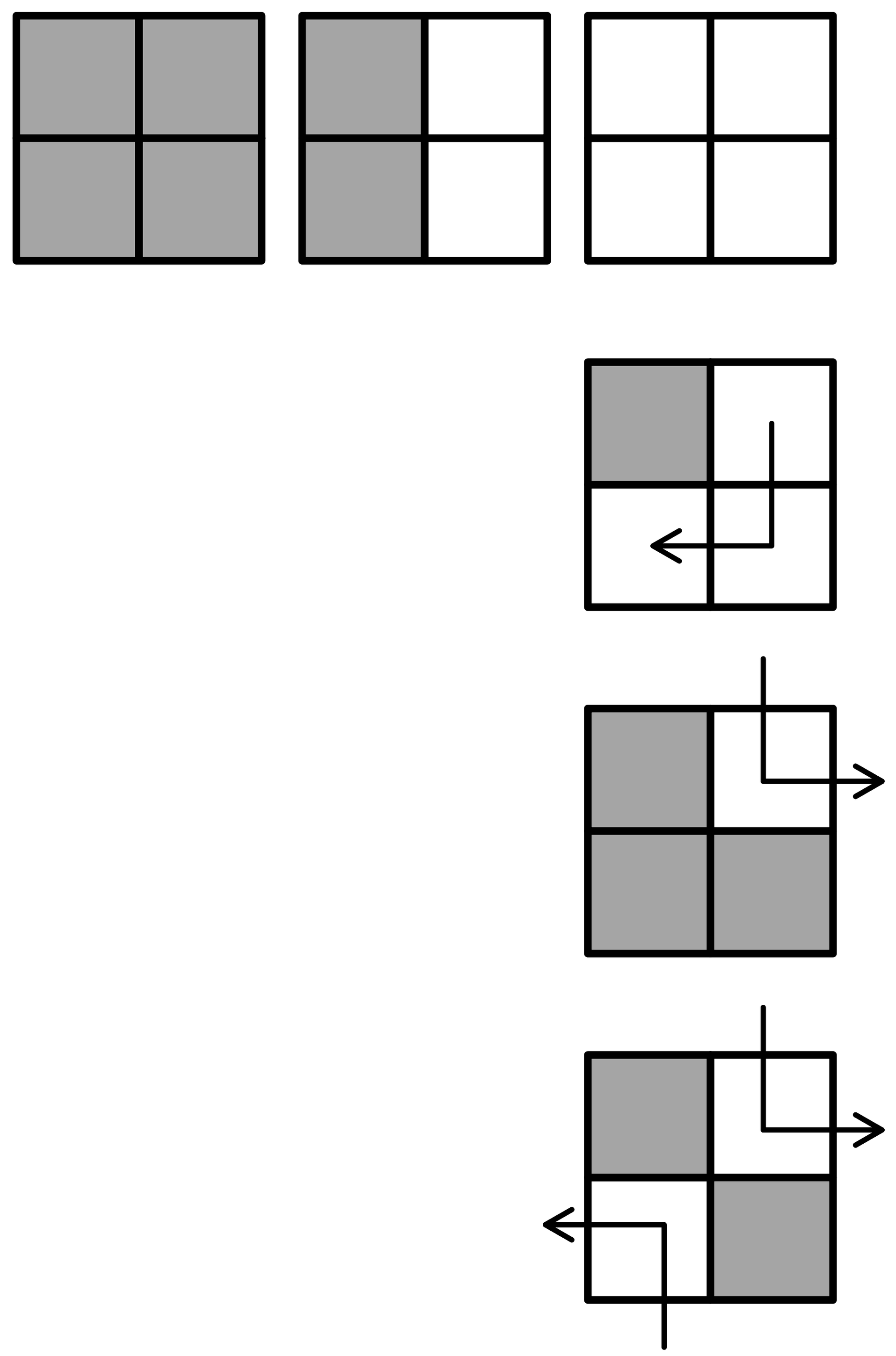


Online clustering



→





0 (no turns)

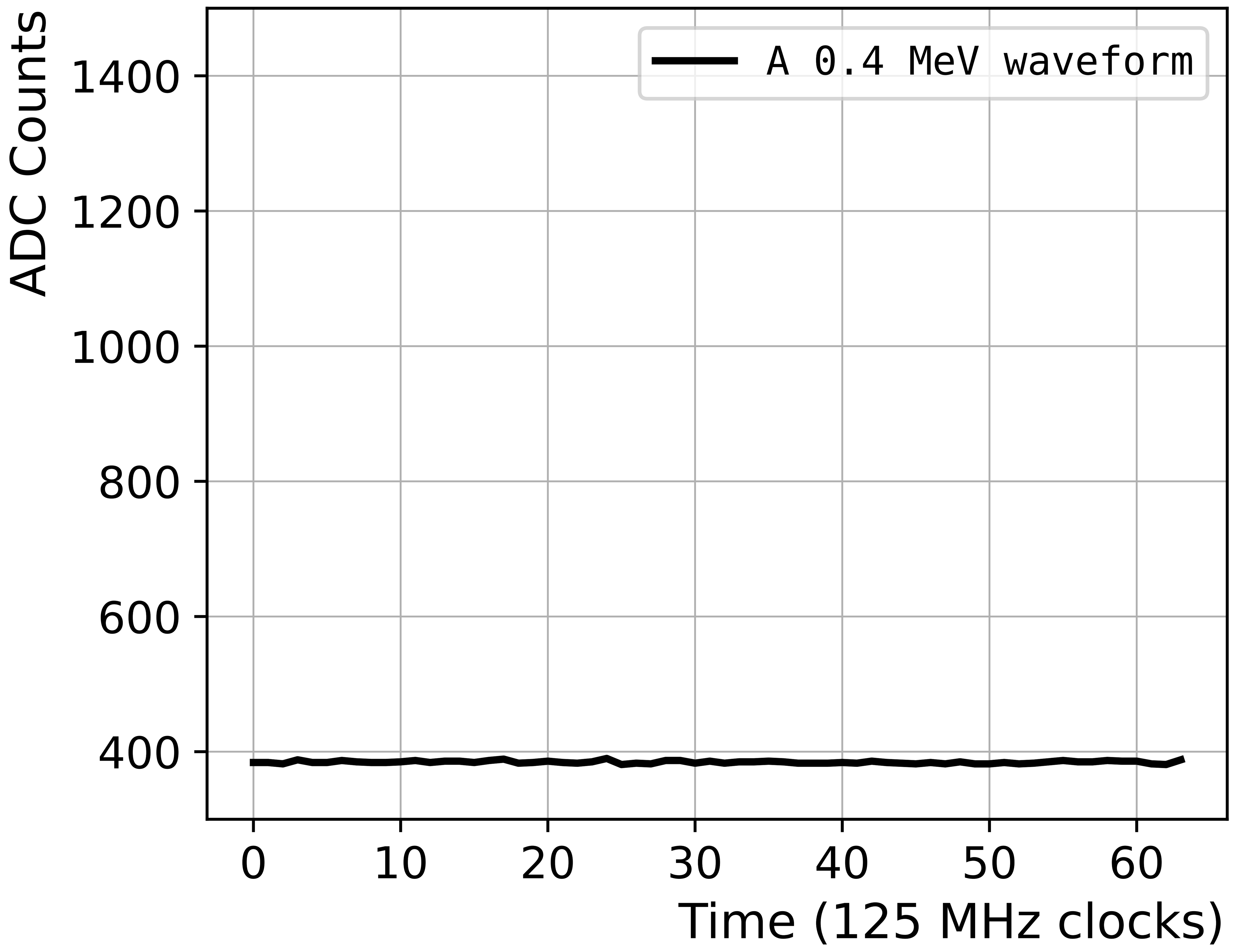
+1 (one turn to the right)

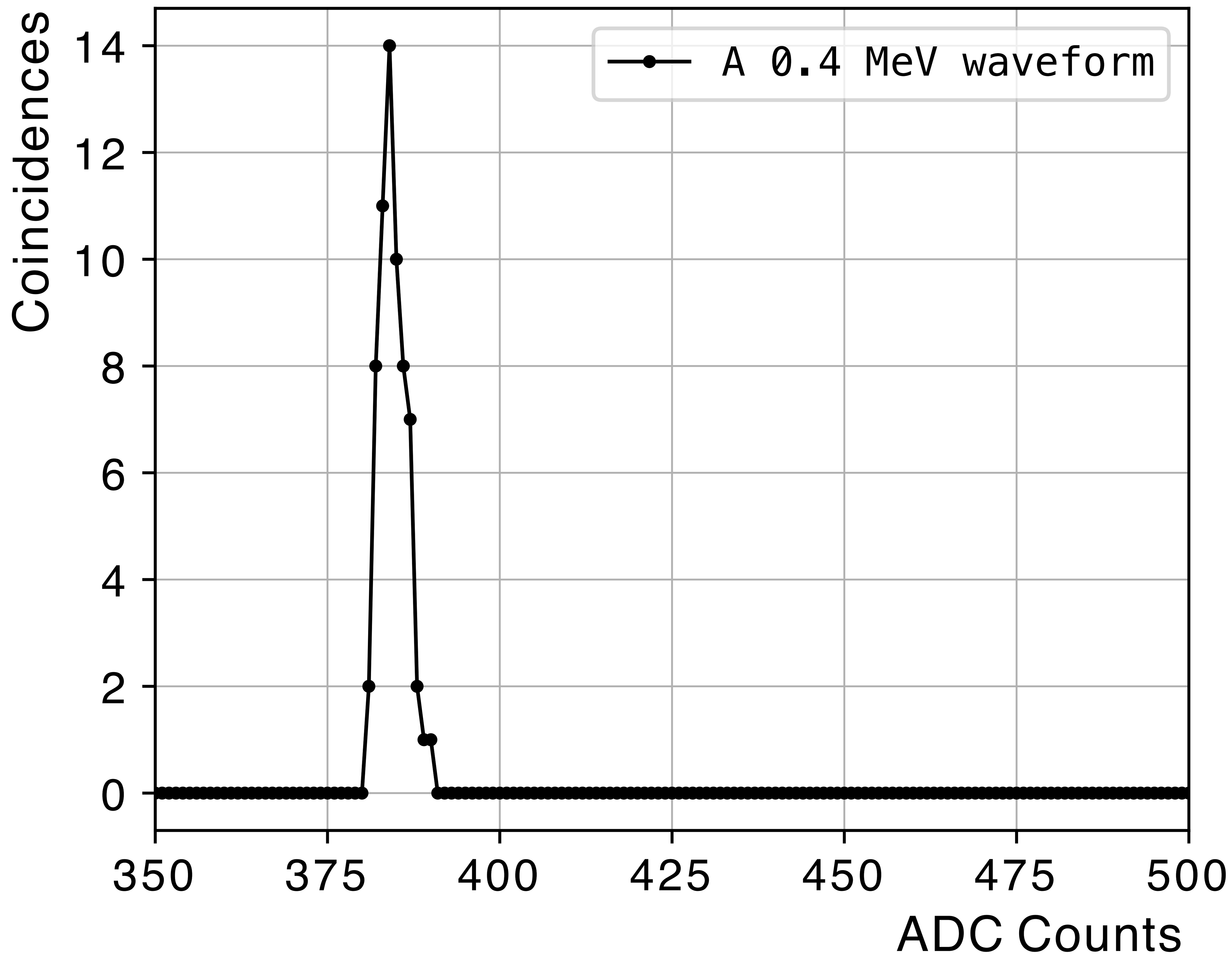
-1 (one turn to the left)

-2 (two turns to the left)

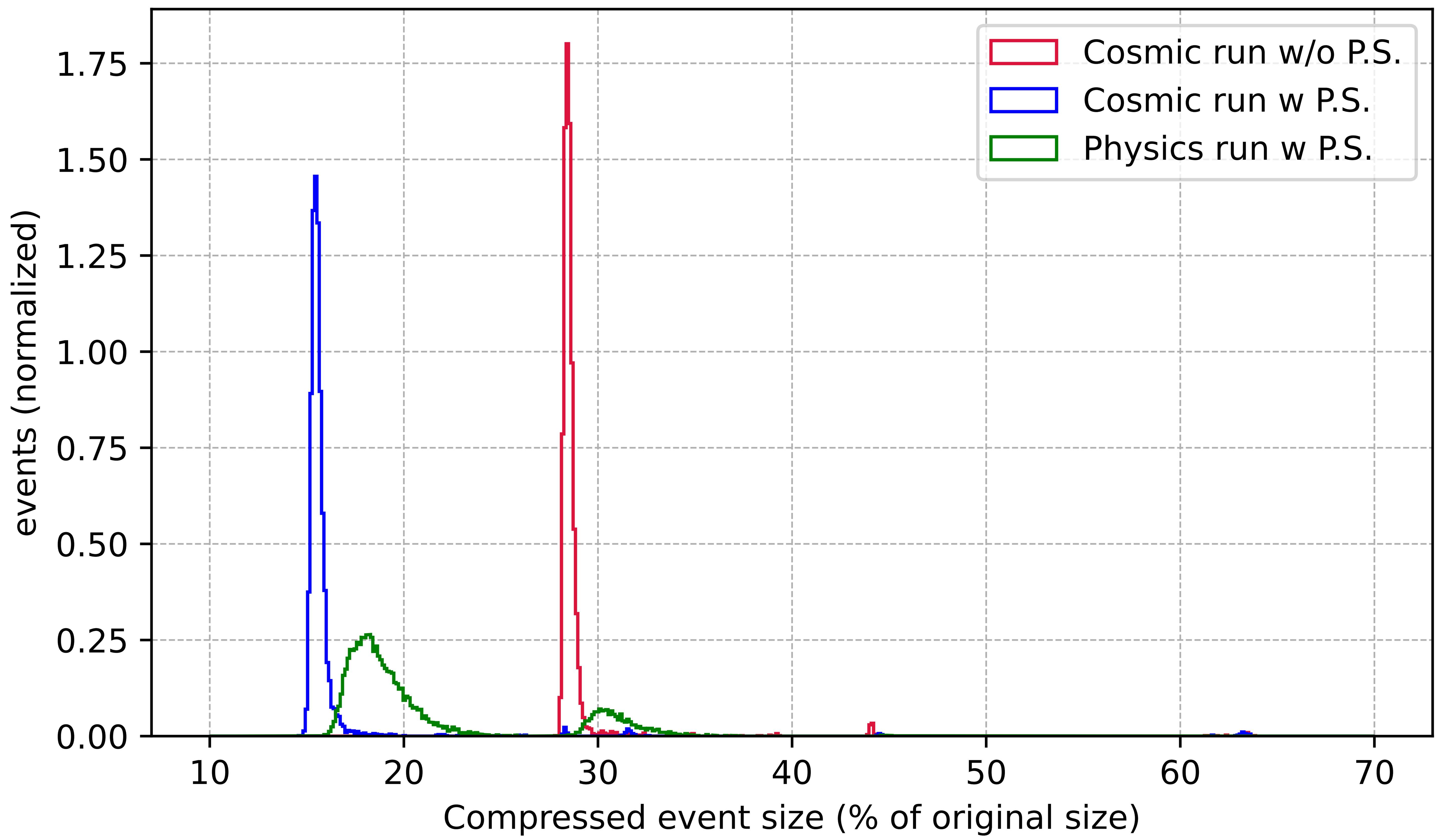
CLUE (the grid)

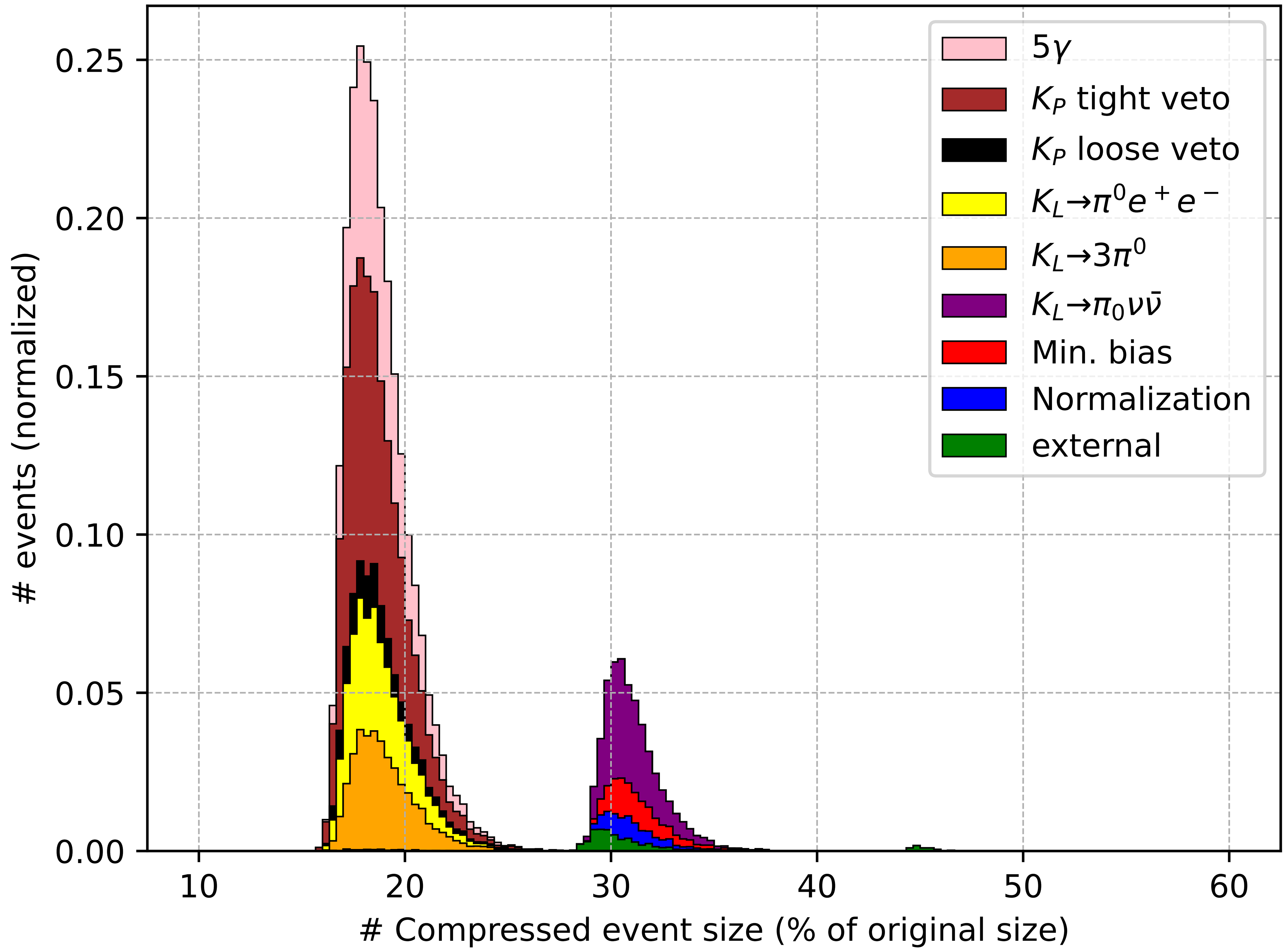
Online pedestal calculation



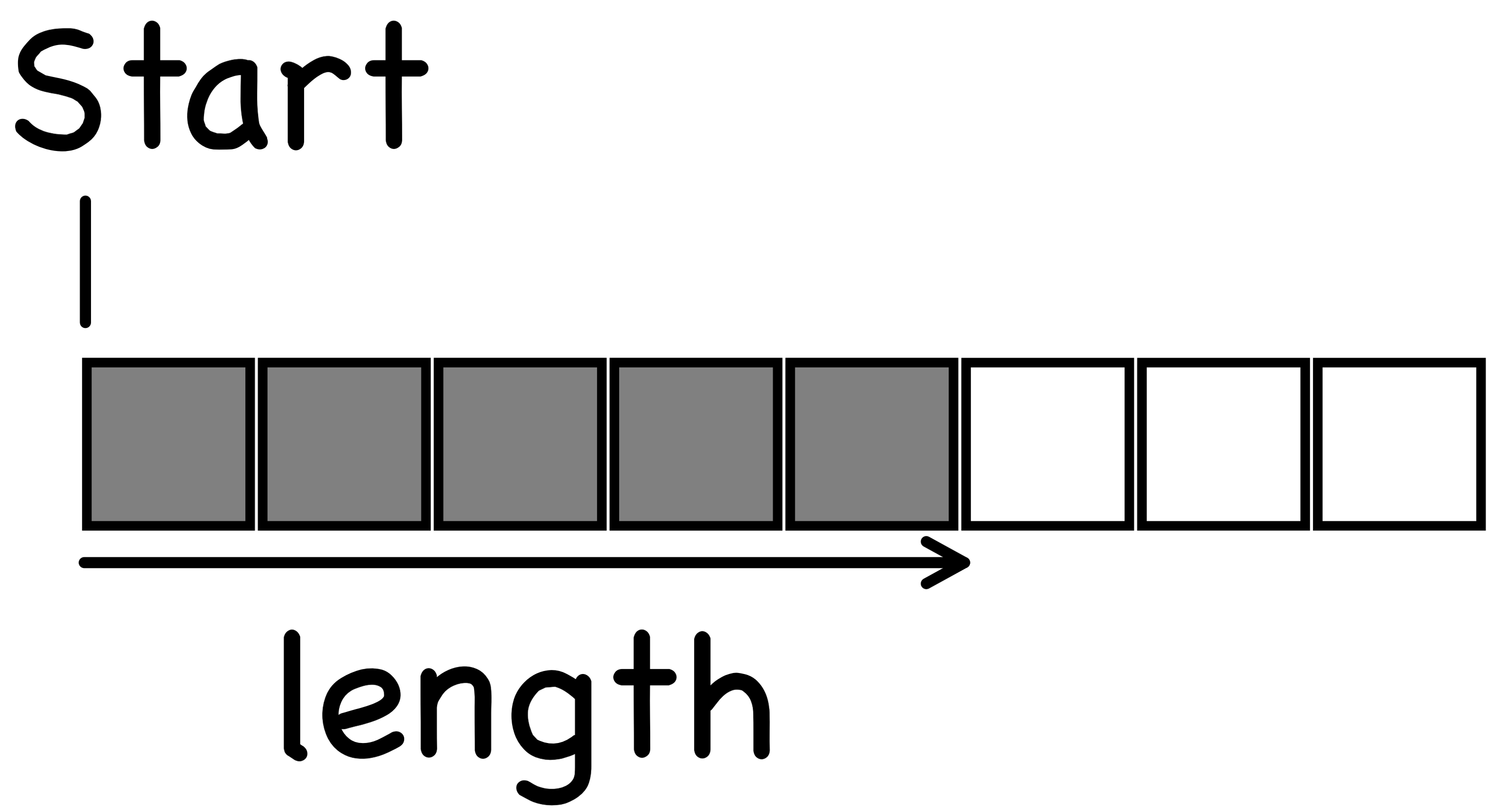


Compression factors

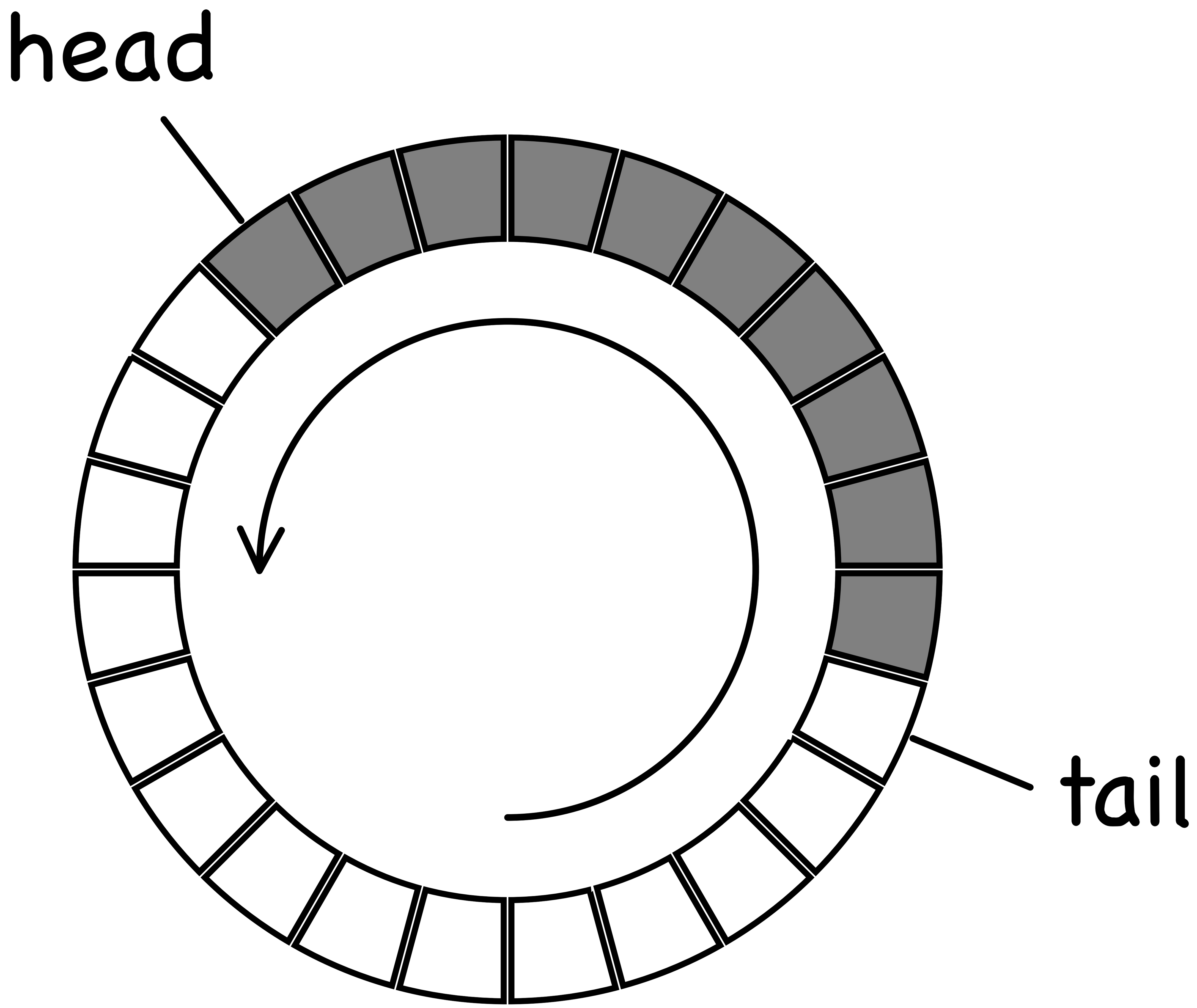




Circular buffers



Linear buffer



Circular buffer